

Variant Identification in *BARD1*, *PRDM9*, *RCC1*, and *RECQL* in Patients with Ovarian Cancer by Targeted Next-generation Sequencing of DNA Pools



Malwina Suszynska¹, Magdalena Ratajska^{2,3}, Paulina Galka-Marciniak¹, Aleksandra Ryszkowska¹, Dariusz Wydra⁴, Jaroslaw Debniak⁴, Anna Jasiak², Bartosz Wasag^{2,5}, Cezary Cybulski⁶, and Piotr Kozlowski¹

ABSTRACT

Several ovarian cancer susceptibility genes have been discovered, but more are likely to exist. In this study, we aimed to analyze knowledge-based selected genes, that is, *BARD1*, *PRDM9*, *RCC1*, and *RECQL*, in which pathogenic germline variants have been reported in patients with breast and/or ovarian cancer. As deep sequencing of DNA samples remains costly, targeted next-generation sequencing of DNA pools was utilized to screen the exons of *BARD1*, *PRDM9*, *RCC1*, and *RECQL* in approximately 400 Polish ovarian cancer cases. A total of 25 pools of 16 samples (including several duplicated samples with known variants) were sequenced on the NovaSeq6000 and analyzed with SureCall (Agilent) application. The set of variants was filtrated to exclude spurious variants, and, subsequently, the identified rare genetic variants were validated using Sanger sequencing. No pathogenic mutation was found within the analyzed cohort of patients with ovarian cancer. Validation genotyp-

ing of filtered rare silent and missense variants revealed that the majority of them were true alterations, especially those with a higher mutation quality value. The high concordance ($R^2 = 0.95$) of population allele frequency for 44 common SNPs in the European control population (gnomAD) and our experiment confirmed the reliability of pooled sequencing. Mutations in *BARD1*, *PRDM9*, *RCC1*, and *RECQL* do not contribute substantially to the risk of ovarian cancer. Pooled DNA sequencing is a cost-effective and reliable method for the initial screening of candidate genes; however, it still requires validation of identified rare variants.

Prevention Relevance: *BARD1*, *PRDM9*, *RCC1*, and *RECQL* are not high/moderate-risk ovarian cancer susceptibility genes. Pooled sequencing is a reliable and cost-effective method to detect rare variants in candidate genes.

Introduction

A substantial fraction of the genetic component involved in ovarian cancer susceptibility remains to be discovered. In this study, we aimed to analyze knowledge-based selected candidate susceptibility genes, that is, *PRDM9*, *RCC1*, and *RECQL*. In

addition, we included in the analysis the gene of our interest, that is, *BARD1*, to experimentally verify our recent observation based on a large-scale case-control association study, showing that *BARD1* is not an ovarian cancer risk gene (1).

PRDM9 (*PR domain zinc finger protein 9*) gene is located on chromosome 5 and codes a zinc finger (ZF) protein with histone methyltransferase activity that binds to chromatin, directing the sites of and inducing DNA double-strand breaks associated with meiotic recombination hotspots (2, 3). *PRDM9* ZF variation were suggested to be a risk factor for chromosomal translocations (4, 5) and a potential reason for acquiring aneuploidies or genomic rearrangements associated with childhood leukemogenesis (6, 7). Most importantly, germline pathogenic mutations in *PRDM9* have recently been reported in patients with ovarian cancer and breast cancer from The Cancer Genome Atlas (TCGA) cohort (8). In addition, although *PRDM9* is expressed predominantly in germ cells and its role in somatic tissues remains unclear, *PRDM9* is expressed in 20% of pan-cancer samples with the high proportion of tumors exhibiting the expressions in ovarian cancer (9).

RECQL (*RecQ Like Helicase*, also known as *RECQL1*) is located on chromosome 12 and encodes an ATP-dependent DNA helicase that is a member of the RecQ DNA helicase

¹Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. ²Department of Biology and Medical Genetics, Medical University of Gdansk, Gdansk, Poland. ³Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand. ⁴Department of Gynecology, Gynecologic Oncology and Gynecologic Endocrinology, Medical University of Gdansk, Gdansk, Poland. ⁵Laboratory of Clinical Genetics, University Clinical Centre, Gdansk, Poland. ⁶Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland.

Note: Supplementary data for this article are available at Cancer Prevention Research Online (<http://cancerprevres.aacrjournals.org/>).

Corresponding Author: Piotr Kozlowski, Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan 61-704, Poland. Phone: 486-1852-8503, ext. 1261; Fax: 486-1852-0532; E-mail: kozlowp@ibch.poznan.pl

Cancer Prev Res 2022;15:151-60

doi: 10.1158/1940-6207.CAPR-21-0295

©2021 American Association for Cancer Research

family (10, 11). RecQ helicases have been shown to be important in genome maintenance, including various types of DNA repair, replication, recombination, and transcription. Mutations in three of five known human RecQ genes are implicated in heritable human disorders: (i) *WRN* mutations in Werner syndrome, (ii) *BLM* mutations in Bloom syndrome, (iii) and *RECQ4* mutations in Rothmund-Thomson syndrome, characterized by an elevated risk for the development of various cancers (12–14). Recently, two independent studies from Poland and China provided convincing evidence that mutations in *RECQL* are associated with breast cancer susceptibility (15, 16). Later, an Australian case–control study did not confirm the role of *RECQL* in breast cancer predisposition (17). As breast cancer and ovarian cancer share some molecular background (most breast cancer predisposition genes to a different extent also predispose to ovarian cancer), it is a priority to determine the role of *RECQL* in ovarian cancer. The inclusion of *RECQL* was further supported by recently identified *RECQL* germline mutations in ovarian cancer cases from TCGA cohort (8).

Finally, *RCC1* is a gene located on chromosome 1 encoding *RCC1* (*regulator of chromosome condensation 1*) protein that binds to chromatin and acts as a guanine exchange factor for Ran (*Ras*-related nuclear protein; ref. 18). *RCC1* is a crucial cell-cycle regulator, and together with other factors in the signal transmission pathway, plays a role in the detection of unrepliated DNA and prevents abnormal chromosomes from entering cell division (19, 20). Methylation-related silencing of *RCC1* expression was related to tumorigenesis and deep invasion in gastric cancer (21). Recently, 19-bp frameshift deletion in exon 10 of *RCC1* was identified in 6 of 159 (3.8%) patients with breast cancer, while no mutation was found among 400 controls. In addition, the mutation partially cosegregates with the disease in the pedigree (22). This small case–control association study suggests that mutations in *RCC1* may be breast cancer high-risk factors; however, *RCC1* has not been tested in ovarian cancer.

Although next-generation sequencing (NGS) technology significantly reduces the cost of per-base sequencing in recent years, the analysis of a large number of samples is still economically challenging. Assuming the cost of \$300 for target capture, library preparation, and sequencing per sample, the cost of a whole project with 400 individuals would be \$120,000, which is limiting for many preliminary studies. A significant part of this amount is the cost of library preparation, which only slightly depends on the size of a region of interest, whether it is several hundred or only four genes. Taking advantage of deep sequencing to reduce the cost and efforts of DNA libraries preparation in our experiment, we opted for an alternative approach based on pooling DNA from several individuals into a larger pool, which is sequenced as a single sample. The approach has been used both for population genetic studies [allele frequency (AF) estimation] and to identify rare mutations associated with different diseases (23–26). Mutation analysis using this approach seems economically justified,

especially for the initial screening of candidate genes where a few mutations can be expected at best.

Therefore, in this study, to elucidate the role of *BARD1*, *PRDM9*, *RCC1*, and *RECQL* in ovarian cancer susceptibility, we analyzed the complete coding sequence of the above genes in 393 patients with ovarian cancer, sequencing a total of 25 pools. Several rare variants, identified in the tested genes by pooled sequencing, were confirmed by an alternative method, but none was a pathogenic mutation.

Materials and Methods

Patients/DNA samples

The study included DNA samples isolated from the whole blood of 393 patients with ovarian cancer who were unselected for age and familial history of the disease. On the basis of the binomial distribution analysis, the number of samples gave us the power of >98% and >85% to detect at least one mutation in a gene, assuming an overall frequency of mutations in the gene 1% and 0.5%, respectively. The samples were collected at the Medical University of Gdansk (Gdansk, Poland) between 1999 and 2012 and were used in previous studies (27–29). In addition, three positive-control DNA samples from patients with breast cancer with known *RECQL* mutation were obtained from the Pomeranian Medical University in Szczecin. The study was approved by the medical review board of the Medical University of Gdansk (Gdansk, Poland; NKBBN/660/2019–2020) and the Pomeranian Medical University in Szczecin (BN-001/33/04) and written informed consent was obtained from all patients, in accordance with the Declaration of Helsinki.

DNA pooling

DNA samples were pooled in equal amounts, in a configuration of 16 individual samples/pool, in a total of 25 pools. Four positive-control samples with three different predefined rare variants in either *BARD1* or *RECQL*, were added to eight pools. One DNA sample with two cooccurring *BARD1* variants (c.609A>C and c.2212A>G) was added to three pools, and three DNA samples with the same *RECQL* mutation (c.1667_1667+3del) were added to five pools.

Targeted enrichment and sequencing

The custom capture probes were designed using the SureDesign eArray portal (Agilent), using the following parameters: tiling density: 2×, masking: moderately stringent, boosting: balanced. The target sequences included all exons (plus 25 nt of upstream and downstream flanking sequences) of the candidate genes, *BARD1*, *PRDM9*, *RCC1*, and *RECQL*, according to the RefSeq (RRID:SCR_003496) annotated transcripts (GRCh37/hg19: NM_000465, NM_020227, NM_001048194, and NM_002907), respectively. In total, the final design covered 52 exons/regions (18,609 kbp) and consisted of 309 probes (21,232 kbp), reaching 100% coverage (exceptions are three a few nucleotide long fragments of untranslated regions). Target capture was performed according to the manufacturer's

protocol (SureSelectXT Low Input Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library, Agilent Technologies; RRID:SCR_013575), employing the unique molecular barcodes (MBC) when detecting very low AF variants. It allows very high coverage sequencing without excessive removal of false duplicates. In brief, unique MBCs are incorporated into each DNA molecule in a sample, allowing distinguishing true (with the same MBC) and false (with different MBCs) duplicates. Paired-end sequencing (2×150 bp read lengths) was performed on the NovaSeq6000 Sequencing System (Illumina), assuming the generation of 3 Gbps of data per pool, to achieve high coverage. Target capture and sequencing were performed in MacroGen (RRID:SCR_014454).

Computational analyses

SureCall (Agilent) was used for end-to-end NGS data analysis. "Single sample" type of analysis was performed for each pool using fastq.gz as input files. For pre-alignment, alignment, and post-alignment steps, the default parameters of SureCall were used. Subsequently, as not to lose any potential variant, the SNPPEP caller was run with less stringent parameters, including minimum alternative allele reads fraction (AARF) of 0.001 and a minimum number of reads supporting variant allele of 1. However, these parameters were raised during subsequent steps of variant filtration, removing spurious variants. Details of the parameters used for the SureCall pipeline are provided in Supplementary Table S1.

Variant filtration and rare variant validation

The following filters for potential pathogenic rare variant were applied: location in the coding region (± 2 bases), AARF ≥ 0.01 , ≥ 5 reads supporting an alternative variant, variant occurrence in ≤ 2 pools.

The variants fulfilling the above criteria were used for validation. The validation excluded variants in exon 11 of *PRDM9* due to sequence complexity/duplications known to generate sequencing errors and precluding proper PCR primer design. To validate/identify a carrier of a candidate variant, all individual samples of the pool in which the variant was detected were PCR amplified and Sanger sequenced. PCR followed the GoTaqG2 DNA Polymerase protocol (Promega), and Sanger sequencing was performed on an ABI Prism 3130 genetic analyzer (Applied Biosystems) according to the manufacturer's general recommendations. The primer sets used for validation of identified variants are shown in Supplementary Table S2.

Comparison of variant frequency between pooled sequencing and population controls

Population AF of common *BARD1*, *PRDM9*, *RCC1*, and *RECQL* variants [i.e., with AF ≥ 0.001 (0.1%) in non-cancer European (non-Finnish) cohort from gnomAD database (30)] was compared with the AF in our cohort, and correlation coefficient R^2 was calculated. The AF of each common variant in the cohort was calculated as a sum of AARFs of the variant in

all pools divided by 0.03125 (i.e., the expected AARF of a single heterozygous variant in the pool of 16 samples; $1/2/16 = 0.03125$), and subsequently divided by 800 (i.e., the sum of all analyzed alleles).

In silico analysis of identified variants

To predict the potential consequence of the identified variants, we use several computational tools, among others, taking into account evolutionary conservation of an amino acid or nucleotide, the location within the protein sequence, and the biochemical consequence of amino acid change [MutationTaster, RRID:SCR_010777 (31), PolyPhen-2, RRID:SCR_013189 (32), SIFT, RRID:SCR_012813 (33), SNPs&GO, RRID:SCR_005788 (34), PROVEAN, RRID:SCR_002182 (35), PANTHER, RRID:SCR_004869 (36), FATHMM (37), MutPred, RRID:SCR_010778 (38)]. For each variant, we calculated the overall damaging score, that is, the sum of damaging scores predicted by each of the algorithms.

Data availability statement

All meaningful data are included in the study. Because all analyses were performed on uncoded pooled NGS data, it is impossible to extract data from individual ovarian cancer samples. This makes the data virtually useless for reanalysis. Therefore, we decided not to deposit the NGS files in a public repository, but the uncoded pooled NGS data are available to share upon request from the corresponding author.

Results

Overall study layout and metrics of pooled NGS

To investigate the role of germline mutations in *BARD1*, *PRDM9*, *RCC1*, and *RECQL* in predisposition to ovarian cancer, we sequenced 393 patients with ovarian cancer who were unselected for age and familial history of the disease, using a pooled sequencing approach. Twenty-five pools were created, each consisting of 16 individual DNA samples, including seven (additional/duplicated) positive-control samples with known rare variants in either *BARD1* or *RECQL*, used for validation purposes. The study design included all exons (plus 25-nt flanks) of the analyzed genes, which translated to 52 target regions (18,609 kb), enriched by 309 bait-probes (altogether overlapping 21,232 kb). The desktop SureCall application was used for variant calling, followed by additional filtration of the obtained variants.

The sequencing was performed on the NovaSeq6000 Sequencing System (Illumina) using 150 bp paired-end reads. On average, 5.2 Gbps of data per pooled sample were generated, with 97.3% and 93.3% of bases reaching a quality score of 20 (Q20) and 30 (Q30), respectively (Supplementary Fig. S1A and S1B). An average of 34.2 M reads per pooled sample were obtained, of which 23.3 M reads passing mapping quality filters were mapped on the human genome using BWA MEM aligner, after duplicate removal [average 8.0 M reads (21.1% of original reads) per pooled sample] (Supplementary Fig. S1C and S1D). Although, as expected, due to the small size of the target

(18.6 kbp), only small fractions of reads [on average, only 336 (1.9%) thousand reads per pooled sample] were mapped to target regions (Supplementary Fig. S1E). The mean coverage depth (per pool) in the target was 1,644 \times , translating to an average coverage of 102 \times (1,664 \times /16) per individual sample in the pool (Supplementary Fig. S1F) and corresponding to the 91.3% and 80.8% of target regions covered at least 500 \times (31 \times per individual sample) and 1,000 \times (63 \times), respectively, which was sufficient for the experiment (Supplementary Fig. S1G).

Variant calling and filtering

In total, 93,035 variants were identified. As was to be expected, due to the assumed very low minimum threshold of AARF at 0.001, the number included a very high proportion of false-positive calls. The variants were then subsequently filtered out using several knowledge-based thresholds/criteria shown in the filtering flow diagram (Supplementary Fig. S2) and described below.

First, we narrowed down the set of variants to variants present only in the coding regions/exons ± 2 bases of the genes, obtaining 21,170 variants. Second, we filtered out all variants with AARF < 0.01 (three times lower than expected ~ 0.03 for heterozygous variant present in one sample in a pool; see Materials and Methods), reducing the list of variants to 1,805. This threshold was determined experimentally, based on analysis of the predetermined positive-control variants ($n = 11$) added to some pools. As shown in Supplementary Table S3, the highest AARF was obtained for the deletion c.1667_1667+3del in *RECQL* in pool 12 and was twice as high as expected (0.064), and the lowest AARF was obtained for the missense variant c.2212A>G in *BARD1* in pool 23 and was approximately 2 \times lower than expected (0.014). Third, we filtered out 134 (7%) variants confirmed by < 5 reads, reducing the number of variants to 1,671. The fraction of these variants may correspond to the fraction of regions not sufficiently covered in our experiment. Finally, we excluded 1,647 variants recurring in more than two pools, considering them either too common to be causative variants or false-positive artifacts of sequencing procedure. The remaining 24 variants were considered rare variants and were subjected for further consideration/validation.

Variant validation

Of the 24 rare variants, 23 were substitutions, and one a dinucleotide insertion. To validate variants and identify the actual carriers, all DNA samples of a pool carrying the variant were resequenced by Sanger sequencing. The analysis confirmed and identified carriers of 15 (62.5%) variants (Fig. 1A, right side), all of them were silent or missense variants. As shown in the inset in Fig. 1A, the not confirmed (false-positive) variants had, on average, much lower mutation quality value (MQV) than true variants (63.0 vs. 118.9, respectively; t test: $P = 0.017$), with 6 of 9 false-positive variants having very low MQV (< 50). Filtering out low MQV variants may allow further

reduction of the false-positive rate, although this level of false-positive rate is acceptable in an analysis of pooled DNA sequencing to avoid the potential loss of true variants. The detailed characteristic of true and false-positive variants is shown in Table 1.

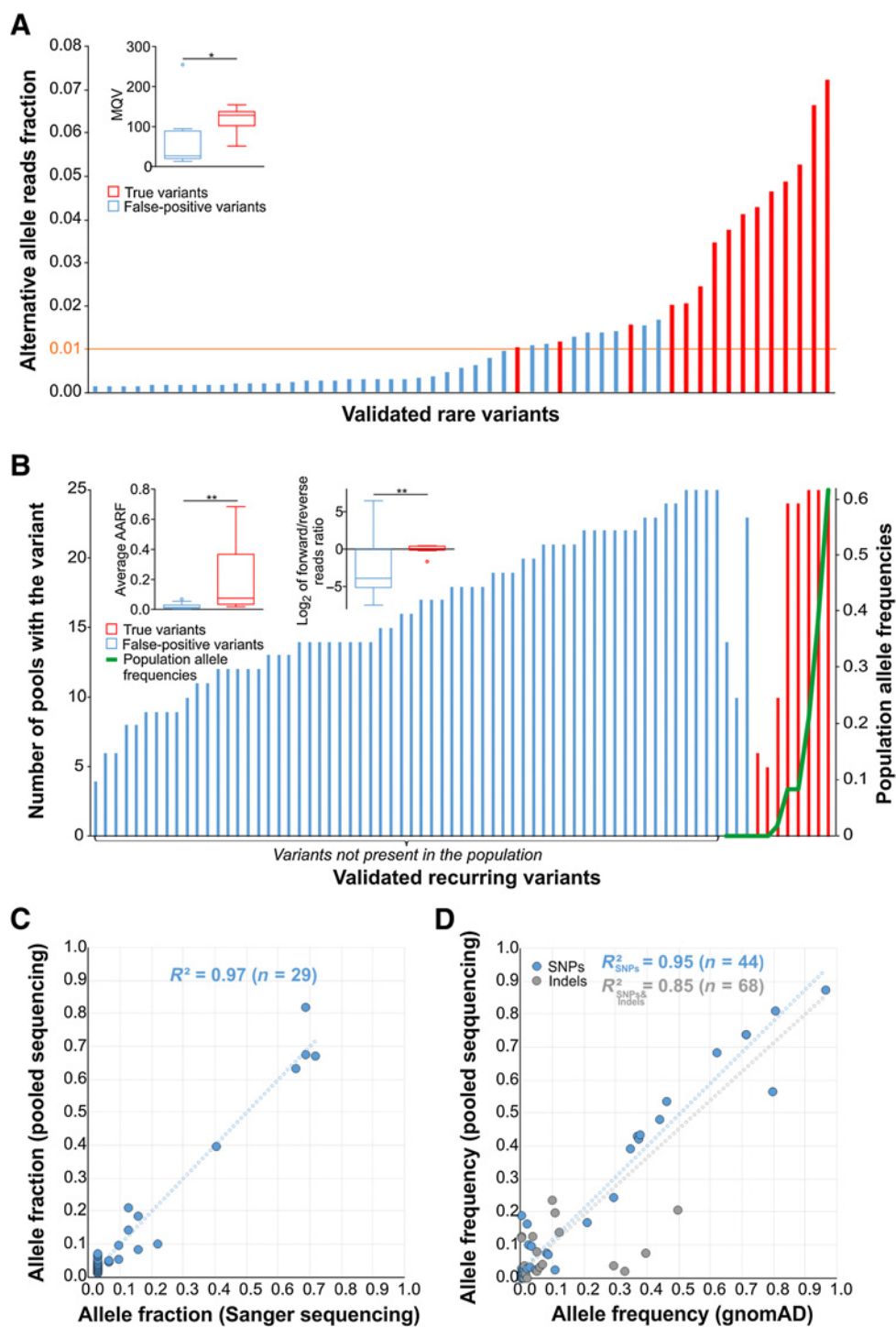
To confirm that we did not lose any variant, we resequenced 29 mutations with AARF lower than the established threshold (i.e., 0.01), mainly focusing on potentially pathogenic variants (i.e., nonsense, frameshift, or splice site variants). None of these variants was confirmed (Fig. 1A, left side), confirming that the assumed AARF threshold at 0.01 is safe for not losing true variants.

For the purpose of evaluation/characterization of the used sequencing strategy, we also validated selected variants recurring in three or more pools (Fig. 1B), considered to be either sequencing artifacts or common/recurring variants. Of the analyzed 73 variants, eight were confirmed as recurrent variants, and all of them were present in the control gnomAD population, whereas the remaining turned out to be false-positive sequencing artifacts and were either (mostly) absent or present at extremely low frequency in gnomAD. As shown in insets in Fig. 1B, the true variants (comparing with false-positive variants) had a significantly higher average AARF (0.191 vs. 0.020, respectively; t test: $P < 0.0001$) and much smaller strand bias [forward to reverse strand ratio; (F test to compare variances: $P < 0.0001$)]. Taking into account the fact that all identified variants that are recurrent in gnomAD are true variants, in total in our cohort, we identified 260 recurring variants (occurring in at least three pools and recurrent in the gnomAD database), including 159 in *BARD1* (9 unique), 9 in *PRDM9* (2 unique), 39 in *RCC1* (2 unique), and 53 in *RECQL* (4 unique). As shown in Supplementary Fig. S3, AF of the recurring variants that passed the population frequency filter is on average higher than AF of the filtered-out variants. Also, the variants that passed filters showed much higher correlation between AF and the number of pools in which they occurred. The characteristic of recurring variants with calculated AF in the analyzed cohort is presented in Table 2 and Supplementary Table S4.

To further evaluate the reliability of our approach for all validated and positive-control variants, we compared the actual allele fractions (based on Sanger sequencing) with AARF (based on pooled sequencing). As shown in Fig. 1C, both values are strongly correlated (0.97). We also compared the frequency of recurring variants determined by pooled sequencing with AF in European population controls ($n = \sim 100,000$) in gnomAD. For the comparison, we selected 68 variants (44 substitutions and 24 indels) present in the control population with a frequency of at least 0.001, which value roughly corresponds to an expected single alternative allele in our cohort of 400 samples. Despite the fact that our cohort (polish population) is not perfectly represented by the general European gnomAD population, the comparison showed a very high concordance for the substitution variants ($R^2 = 0.95$; Fig. 1D; Supplementary Table S5) and was only

Figure 1.

Validation of pool-sequencing approach. **A**, Validation of rare variants (identified in one or two pools) by Sanger sequencing. The variants with $AARF \geq 1$ were selected during the filtration steps. Additional variants are randomly selected loss-of-function mutations with $AARF \leq 0.01$. Inset, Comparison of MQV between false-positive variants and true variants. **B**, Validation of variants identified in three or more pools. Each bar represents a different recurring variant. Red bars indicated true recurring variants, while blue bars indicate false-positive recurrent variants. All true recurrent variants were present in population controls. Inset (left), Comparison of average AARF between false-positive variants and true variants. Inset (right), Comparison of forward/reverse alternative allele reads ratio between false-positive and true variants. **C**, Comparison of allele fraction for confirmed variants between pooled sequencing and Sanger sequencing. **D**, Comparison of AF between pooled sequencing and gnomAD database, for single-nucleotide variants ($n = 44$; blue dots) and all variants ($n = 88$), including deletions and insertions (gray dots).



slightly lower when indel variants were included ($R^2 = 0.85$). The lower concordance for the indel variants most likely results from the fact that a large proportion of indels were multiallelic variants, located in a stretch of low-complexity sequences that still represent a challenge for NGS technology/algorithms and may generate artifacts both in our analysis and in the database.

Characteristic of identified true rare variants

In total, 15 true rare variants were detected in the four analyzed genes in 15/393 (3.8%) patients with ovarian cancer. Of the 15 variants, 6 were found in *BARD1* (5 unique), 6 in *PRDM9* (5 unique), 2 in *RCC1* (1 unique), and 1 in *RECQL*. Six variants (5 unique) and 9 variants (7 unique) were silent and missense variants, respectively. Most of the

Downloaded from <http://aacrjournals.org/cancerpreventionresearch/article-pdf/15/3/151/3051819/151.pdf> by guest on 09 December 2023

Table 1. Characteristic of rare variants selected for validation with Sanger sequencing.

Position	Transcript; protein consequence	dbSNP ID	No. of pools with the variant	MQV	AARF (number of variant alleles/read depth)	Sanger confirmation	Allele frequency from gnomAD (number of variant alleles/number of all alleles)
RCC1 (NM_001048194)							
chr1:28862396C>T (EX8)	c.768C>T; p.Val256=	rs145520116	2	137	0.0463 (283/6,113)	Confirmed	0.00044 (51/115,218)
chr1:28862396C>T (EX8)	c.768C>T; p.Val256=	rs145520116	2	135	0.0486 (167/3,435)	Confirmed	0.00044 (51/115,218)
BARD1 (NM_000465)							
chr2:215593452C>T (EX11)	c.2282G>A; p.Ser761Asn	rs142155101	2	128	0.0376 (149/3,961)	Confirmed	0.00153 (181/117,962)
chr2:215593452C>T (EX11)	c.2282G>A; p.Ser761Asn	rs142155101	2	142	0.0526 (53/1,007)	Confirmed	0.00153 (181/117,962)
chr2:215593483G>A (EX11)	c.2251C>T; p.Arg751Trp	rs139785364	1	110	0.0244 (34/1,393)	Confirmed	0.00001 (1/117,986)
chr2:215595164G>A (EX10)	c.1972C>T; p.Arg658Cys	rs3738888	1	94	0.0130 (7/538)	—	0.00842 (994/118,078)
chr2:215633990G>A (EX5)	c.1361C>T; p.Pro454Leu	rs772285343	1	148	0.0722 (50/693)	Confirmed	—
chr2:215645755T>C (EX4)	c.843A>G; p.Pro281=	rs1559424812	1	90	0.0116 (37/3,179)	Confirmed	—
chr2:215674261C>A (EX1)	c.33G>T; p.Gln11His	rs143914387	1	154	0.0663 (99/1,494)	Confirmed	0.00189 (171/90,716)
PRDM9 (NM_020227)							
chr5:23510127A>T (EX4)	c.292A>T; p.Arg98Trp		1	20	0.0138 (18/1,308)	—	—
chr5:23521145C>A (EX6)	c.365C>A; p.Alal22Glu	rs773183075	2	120	0.0350 (74/2,144)	Confirmed	0.00002 (6/117,484)
chr5:23521145C>A (EX6)	c.365C>A; p.Alal22Glu	rs773183075	2	129	0.0410 (40/975)	Confirmed	0.00002 (6/117,484)
chr5:23521146G>A (EX6)	c.366G>A; p.Alal22=	rs766083673	1	107	0.0202 (49/2,426)	Confirmed	0.00004 (5/117,468)
chr5:23522825C>T (EX8)	c.713C>T; p.Ser238Leu		1	83	0.0096 (51/5,303)	—	—
chr5:23522988C>T (EX8)	c.876C>T; p.Ser292=	rs376338646	1	131	0.0428 (42/981)	Confirmed	0.00011 (13/117,768)
chr5:23524460G>A (EX10)	c.968G>A; p.Arg323Gln	rs183638311	1	51	0.0103 (27/2,627)	Confirmed	0.00353 (415/117,610)
chr5:23524516A>G (EX10)	c.1024A>G; p.Arg342Gly	rs193211869	1	102	0.0206 (52/2,520)	Confirmed	0.00337 (396/117,610)
RECQL (NM_002907)							
chr12:21623231T>A (EX15)	c.1847A>T; p.Glu616Val		1	13	0.0111 (5/452)	—	—
chr12:21623243T>A (EX15)	c.1835A>T; p.Asp612Val		1	27	0.0139 (6/431)	—	—
chr12:21623281C>CAA (EX15)	c.1798-2_1798-linsAA; -		1	255	0.0109 (12/1,099)	—	—
chr12:21629852C>G (EX8)	c.942G>C; p.Gly314=	rs774097655	1	100	0.0155 (8/515)	Confirmed	0.00008 (8/101,004)
chr12:21639487G>C (EX5)	c.427C>G; p.Leu143Val		1	20	0.0156 (40/2,558)	—	—
chr12:21639509G>C (EX5)	c.405C>G; p.Leu135=	rs747916666	1	26	0.0141 (23/1,634)	—	—
chr12:21639510A>C (EX5)	c.404T>G; p.Leu135Arg		1	29	0.0167 (20/1,195)	—	—

variants were previously reported in population controls (gnomAD) [exceptions constitute two variants in *BARD1* not reported in gnomAD but reported in Clinvar, that is, c.843A>G (p.Pro281=), and c.1361C>T; (p.Pro454Leu)]. *BARD1* variants (except the c.843A>G) were also reported previously in literature in patients with breast cancer and/or ovarian cancer (27, 39–45).

Although none of the variants was definitively deleterious to evaluate the potential effect of the missense variants, we analyzed them with several *in silico* classifiers. As shown in **Table 3**, for three of the variants, that is, c.1361C>T (p.Pro454Leu) in *BARD1*, c.2251C>T (p.Arg751Trp) in *BARD1*, and c.1024A>G (p.Arg342Gly) in *PRDM9*, the overall damaging score was substantially higher (7/8, 6.5/8, and 4.5/8, respectively) than for others. Also, the overall damaging score was higher for the rare missense variants than for the analyzed for comparison recurrent missense variants (3.500 vs. 1.143, respectively; *t* test: *P* = 0.0445). Therefore, although we did not detect any definitive pathogenic mutations supporting the role of the investigated genes in predisposition to ovarian cancer, we cannot exclude that some of the identified rare variants, especially those with the highest damaging score, may be functional variants. Indeed, c.1361C>T has been shown experimentally to cause exon 5 skipping in *BARD1* (in-frame

deletion from c.1315 to c.1395), resulting in disruption of the first and second ankyrin repeat on the protein level (27).

Discussion

In the study, we tested the hypothesis that *BARD1*, *PRDM9*, *RCC1*, and *RECQL* are enriched in pathogenic mutations in patients with ovarian cancer. Because no definitive pathogenic variant was identified in the studied genes, this suggests a lack of association of these genes with the high predisposition to ovarian cancer or that the pathogenic variants in these genes in ovarian cancer are too rare to be detected in the substantial group of 400 cases that allowed the identification of mutation at the occurrence level of 0.25%. The latter would indicate that mutations in these genes are at best low or moderate risk variants. Also, computational analyses demonstrated the damaging potential of a few rare missense variants identified in our study, therefore their clinical relevance cannot be excluded. Further large-scale association studies are required to determine the ovarian cancer risk associated with *PRDM9*, *RCC1*, and *RECQL* mutations including also variants of uncertain significance and common SNPs. Regarding *BARD1*, the current evidence seems to be sufficient to exclude the association of *BARD1* mutations with the ovarian cancer risk.

Table 2. Characteristic of recurring variants.

Position	Transcript; protein consequence	dbSNP ID	No. of pools with the variant	Validation in X samples	Total allele frequency based on AARF	Allele frequency from gnomAD (number of variant alleles/number of all alleles)
RCC1 (NM_001048194)						
chr1:28861636G>A (EX6)	c.609G>A; p.Val203=	rs2066726	25		0.24454	0.29614 (34,275/115,738)
chr1:28864435A>G (EX11)	c.1275A>G; p.Lys425=	rs2227977	14		0.03378	0.02829 (3,263/115,338)
BARD1 (NM_000465)						
chr2:215593522T>C (EX11)	c.2212A>G; p.Ile738Val	rs61754118	7	3 (KV)	0.01213	0.00822 (971/118,058)
chr2:215617178C>G (EX7)	c.1670G>C; p.Cys557Ser	rs28997576	12		0.02967	0.02263 (2,672/118,066)
chr2:215632255C>T (EX6)	c.1519G>A; p.Val507Met	rs2070094	25		0.42280	0.37124 (43,806/117,998)
chr2:215632256A>G (EX6)	c.1518T>C; p.His506=	rs2070093	25		0.81188	0.80515 (95,030/118,028)
chr2:215645464C>G (EX4)	c.1134G>C; p.Arg378Ser	rs2229571	25	4 (VV)	0.68524	0.61841 (72,967/117,992)
chr2:215645503_215645523del	c.1075_1095del;	rs28997575	8	2 (VV)	0.01260	0.01800 (2,125/118,060)
TGGTGAAGAACATTTCAGGCAA (EX4)	p.Leu359_Pro365del					
chr2:215645545C>G (EX4)	c.1053G>C; p.Thr351=	rs2070096	25	4 (VV)	0.16934	0.21188 (25,003/118,006)
chr2:215645989T>G (EX4)	c.609A>C; p.Gly203=	rs28997574	7	3 (KV)	0.01099	0.00958 (1,084/113,106)
chr2:215674224G>A (EX1)	c.70C>T; p.Pro24Ser	rs1048108	25	1 (VV)	0.43316	0.37841 (41,354/109,284)
PRDM9 (NM_020227)						
chr5:23509200C>T (EX2)	c.58C>T; p.Arg20Trp	rs184600328	4	1 (VV)	0.00334	0.00140 (165/117,690)
chr5:23524485G>A (EX10)	c.993G>A; p.Val331=	rs188381798	5	3 (VV)	0.00944	0.00138 (162/117,616)
RECQL (NM_002907)						
chr12:21623179T>C (EX15)	c.1899A>G; p.Gln633=	rs17849408	23	2 (VV)	0.06910	0.08491 (9,899/116,588)
chr12:21623969A>G (EX14)	c.1731T>C; p.Asn577=	rs6500	22	2 (VV)	0.07528	0.08317 (9,504/114,272)
chr12:21624359_21624362delTACT (EX13)	c.1667_1667+3del; p.Glu557Lysfs*14	rs564485792	5	5 (KV)	0.00712	0.00052 (61/116,822)
chr12:21624546C>G (EX13)	c.1483G>C; p.Asp495His	rs6499	3		0.00546	0.00495 (582/117,484)

Abbreviations: KV, known variant (positive-control variant); VV, validated recurring variant.

Another goal of the study was to test the suitability of pooling samples at the level of genomic DNA before library preparation and sequencing for rare variant detection. We were able to detect both common and rare variants, including positive-control variants. Validation with an alternative Sanger sequencing method confirmed that many variants, identified in only one or two pools, were true rare variants. As shown in the validation section, the accuracy of calling rare and recurring variants could be improved with additional filters on MQV, average AARF, AF or forward/reverse reads ratio. In addition, the comparison of variant frequency between pooled sequencing and the population controls (in our study and others) provides evidence of the accuracy of the results from pooled sequencing (25, 46, 47). The approach allowed us to use a smaller amount of DNA, which is extremely important when handling archival DNA samples used in multiple studies. In addition, it saved time and effort invested in the experiment and, more importantly, significantly reduced the cost of the experiment compared with sequencing individual samples with both NGS and traditional Sanger sequencing. Including a small capture (<500 kbp) and a sample size of approximately 400 patients, using the pools of 16 samples allowed to reduce the cost of the experiment by approximately 11× (see Table 4). Lowering the cost of the experiment allows the inclusion of more patients in the preliminary candidate-testing experiments, which increases the power to detect pathogenic mutations or other significant AF differences. Common and rare allele

selection for further association studies or associations directly from the pooled sequencing results has already been performed (23, 46, 48, 49). Performing a targeted NGS of DNA pools, known cancer susceptibility genes (i.e., *BRCA1/2*, *PALB2*) can be added to a gene panel alongside candidate genes to serve as a mutation detection control and to better characterize the tested population. Summarizing, for an initial screening of selected candidate genes, sequencing of DNA pools seems to be a cost, DNA, and effort saving strategy.

However, it should be noted that the pooling strategy requires additional methodologic steps, including (i) careful preparation of the pools (i.e., sample selection and pooling in terms of quantity and quality), and (ii) validation of the variant and determination of the actual carrier in the pool (i.e., sequencing all samples in the pool separately). Finally, the limitation of the pooling strategy is the possibility of overlooking some variants when some regions in the sample are inadequately covered. Therefore, we would like to clearly state that, we do not recommend the pooling for diagnostic purposes but for fast and cost-effective testing of candidate genes with expected rare pathogenic variants.

Concluding, our results show a lack of pathogenic mutations in *BARD1*, *PRDM9*, *RCC1*, and *RECQL* in ovarian cancer, indicating no or low contribution of the analyzed genes to ovarian cancer susceptibility. We also show the advantage of the pooled sequencing approach that is DNA-, an effort-, and a cost-effective method to detect rare variants in candidate genes.

Table 3. Predicted effect of the missense variants identified in patients with ovarian cancer.

Transcript; protein consequence	AF from pooled sequencing (number of variant alleles/number of all alleles)	AF from gnomAD (number of variant alleles/number of all alleles)	ClinVar	MutationTaster	PolyPhen-2	SIFT	SNP&Go	PROVEAN	PANTHER	FATHMM	MutPred	Score
Rare variants												
<i>BARD1</i> ; c.2282G>A; p.Ser761Asn	0.00250 (2/800)	0.00153 (181/117,962)	CIP	1	0	0	1	0	0	0	0	2
<i>BARD1</i> ; c.2251C>T; p.Arg751Trp	0.00125 (1/800)	0.00001 (1/117,986)	US	1	1	1	1	1	1	0	0.5	6.5
<i>BARD1</i> ; c.1361C>T; p.Pro454Leu	0.00125 (1/800)	—	US	1	1	1	1	1	1	0	1	7
<i>BARD1</i> ; c.33G>T; p.Gln11His	0.00125 (1/800)	0.00189 (171/90,716)	B/LB	0	0	0	0	0	0	0	0	0
<i>PRDM9</i> ; c.365C>A; p.Alal22Glu	0.00250 (2/800)	0.00002 (6/117,484)	—	0	0.5	1	0	0	0	0	0	1.5
<i>PRDM9</i> ; c.968G>A; p.Arg323Gln	0.00125 (1/800)	0.00353 (415/117,610)	—	1	1	1	—	0	0	0	0	3
<i>PRDM9</i> ; c.1024A>G; p.Arg342Gly	0.00125 (1/800)	0.00337 (396/117,610)	—	1	1	1	—	1	0	0	0.5	4.5
Recurring variants												
<i>BARD1</i> ; c.2212A>G; p.Ile738Val ^a	0.01213	0.00822 (971/118,058)	B/LB	1	0	0	0	0	0.5	0	0	1.5
<i>BARD1</i> ; c.1670G>C; p.Cys557Ser	0.02967	0.02263 (2,672/118,066)	B	0	0	0	1	1	0.5	0	0	2.5
<i>BARD1</i> ; c.1519G>A; p.Val507Met	0.42280	0.37124 (43,806/117,998)	B/LB	0	0	0	0	0	0	0	0	1
<i>BARD1</i> ; c.1134G>C; p.Arg378Ser	0.68524	0.61841 (72,967/117,992)	B	0	0	0	1	0	0	0	0	1
<i>BARD1</i> ; c.70C>T; p.Pro24Ser	0.43316	0.37841 (41,354/109,284)	B	0	0	0	0	0	0	0	0	0
<i>PRDM9</i> ; c.58C>T; p.Arg20Trp	0.00334	0.00140 (165/117,690)	—	0	0	0	—	0	0	0	0	0
<i>RECQL1</i> ; c.1483G>C; p.Asp495His	0.00546	0.00495 (582/117,484)	B/LB	1	0	0	0	1	—	0	0	2

Abbreviations: ClinVar: B, benign; C, conflicting interpretations of pathogenicity; LB, likely benign; US, uncertain significance; -, not reported in ClinVar; MutationTaster: polymorphism = 0, disease causing = 1; PolyPhen-2: benign = 0, possibly damaging = 0.5, probably damaging = 1; SIFT: tolerated = 0, damaging = 1; SNPs&Go: neutral = 0, disease = 1; PROVEAN: neutral = 0, deleterious = 1; PANTHER: possibly benign = 0, possibly damaging = 0.5, probably damaging = 1; FATHMM: tolerated = 0, damaging = 1; MutPred2: <0.5 = 0, 0.5-0.75 = 0.5, >0.75 = 1.

^aKnown variant in 3/7 pools.

Table 4. Comparison of the cost between pooled and individual sequencing.

	Pooled sequencing	Individual-DNA NGS sequencing	Individual-DNA Sanger sequencing
Library preparation	25pools*272\$ = 6,800\$	393 ^a samples*272\$ = 106,896\$	—
NGS Sequencing	25pools*3Gb*24\$ = 1,800\$	393 ^a samples*1Gb*24\$ = 9,432\$	—
Validation step (PCR+Sanger sequencing)	24variants*16samples*6\$ = 2,304\$	24variants*1sample*6\$ = 144\$	52exons*393 ^a samples*6\$ = 122,616\$
Sum (fold increase)	10,904\$ (-)	116,472\$ (10.7x)	122,616\$ (11.2x)

^aThe exact number of patients with ovarian cancer.

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

M. Suszynska: Conceptualization, formal analysis, validation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **M. Ratajska:** Resources, writing—review and editing. **P. Galka-Marciniak:** Formal analysis, investigation, writing—review and editing. **A. Ryzkowska:** Validation, writing—review and editing. **D. Wydra:** Resources, writing—review and editing. **J. Debniak:** Resources, writing—review and editing. **A. Jasiak:** Resources, writing—review and editing. **B. Wasag:** Resources, writing—review and editing. **C. Cybulski:** Resources, writing—review and editing. **P. Kozłowski:** Conceptualization, resources,

formal analysis, supervision, funding acquisition, investigation, methodology, writing—original draft, writing—review and editing.

Acknowledgments

This work was supported by grants from the Polish National Science Centre (NCN 2015/17/B/NZ2/01182 and 2016/22/A/NZ2/00184) awarded to P. Kozłowski.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 13, 2021; revised October 27, 2021; accepted December 2, 2021; published first December 10, 2021.

References

- Suszynska M, Kozłowski P. Summary of BARD1 mutations and precise estimation of breast and ovarian cancer risks associated with the mutations. *Genes* 2020;11:798.
- Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 2010;327:836–40.
- Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, et al. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* 2010;327:876–9.
- Borel C, Cheung F, Stewart H, Koolen DA, Phillips C, Thomas NS, et al. Evaluation of PRDM9 variation as a risk factor for recurrent genomic disorders and chromosomal non-disjunction. *Hum Genet* 2012;131:1519–24.
- Berg IL, Neumann R, Lam KW, Sarbajna S, Odenthal-Hesse L, May CA, et al. PRDM9 variation strongly influences recombination hotspot activity and meiotic instability in humans. *Nat Genet* 2010;42:859–63.
- Hussin J, Sinnott D, Casals F, Idaghdour Y, Bruat V, Saillour V, et al. Rare allelic forms of PRDM9 associated with childhood leukemogenesis. *Genome Res* 2013;23:419–30.
- Woodward EL, Olsson ML, Johansson B, Paulsson K. Allelic variants of PRDM9 associated with high hyperdiploid childhood acute lymphoblastic leukaemia. *Br J Haematol* 2014;166:947–9.
- Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic germline variants in 10,389 adult cancers. *Cell* 2018;173:355–70.
- Houle AA, Gibling H, Lamaze FC, Edgington HA, Soave D, Fave MJ, et al. Aberrant PRDM9 expression impacts the pan-cancer genomic landscape. *Genome Res* 2018;28:1611–20.
- Puranam KL, Blackshear PJ. Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ. *J Biol Chem* 1994;269:29838–45.
- Seki M, Miyazawa H, Tada S, Yanagisawa J, Yamaoka T, Hoshino S, et al. Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to Escherichia coli Rec Q helicase and localization of the gene at chromosome 12p12. *Nucleic Acids Res* 1994;22:4566–73.
- Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, et al. Positional cloning of the Werner's syndrome gene. *Science* 1996;272:258–62.
- Kitao S, Shimamoto A, Goto M, Miller RW, Smithson WA, Lindor NM, et al. Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. *Nat Genet* 1999;22:82–4.
- Ellis NA, Groden J, Ye TZ, Straughen J, Lennon DJ, Ciocci S, et al. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 1995;83:655–66.
- Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, Wokolorczyk D, et al. Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* 2015;47:643–6.
- Sun J, Wang Y, Xia Y, Xu Y, Ouyang T, Li J, et al. Mutations in RECQL gene are associated with predisposition to breast cancer. *PLoS Genet* 2015;11:e1005228.
- Li N, Rowley SM, Goode DL, Amarasinghe KC, McInerney S, Devereux L, et al. Mutations in RECQL are not associated with breast cancer risk in an Australian population. *Nat Genet* 2018;50:1346–8.
- Bischoff FR, Ponstingl H. Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. *Nature* 1991;354:80–2.
- Dasso M. RCC1 in the cell cycle: the regulator of chromosome condensation takes on new roles. *Trends Biochem Sci* 1993;18:96–101.
- Ren X, Jiang K, Zhang F. The multifaceted roles of RCC1 in tumorigenesis. *Front Mol Biosci* 2020;7:225.
- Lin YL, Chen HL, Cheng SB, Yeh DC, Huang CC, P'Eng FK, et al. Methylation-silencing RCC1 expression is associated with tumorigenesis and depth of invasion in gastric cancer. *Int J Clin Exp Pathol* 2015;8:14257–69.

22. Riahi A, Radmanesh H, Schurmann P, Bogdanova N, Geffers R, Meddeb R, et al. Exome sequencing and case-control analyses identify RCC1 as a candidate breast cancer susceptibility gene. *Int J Cancer* 2018;142:2512–7.
23. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;43:1066–73.
24. Calvo SE, Tucker EJ, Compton AG, Kirby DM, Crawford G, Burtt NP, et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet* 2010;42:851–8.
25. Anand S, Mangano E, Barizzone N, Bordoni R, Sorosina M, Clarelli F, et al. Next generation sequencing of pooled samples: guideline for variants' filtering. *Sci Rep* 2016;6:33735.
26. Pihlstrom L, Rengmark A, Bjornara KA, Toft M. Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease. *Ann Hum Genet* 2014;78:243–52.
27. Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, et al. Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms. *Oncol Rep* 2015;34:2609–17.
28. Klonowska K, Ratajska M, Czubak K, Kuzniacka A, Brozek I, Koczkowska M, et al. Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example. *Sci Rep* 2015;5:10424.
29. Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, et al. The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population. *Oncotarget* 2017;8:76357–74.
30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
31. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11:361–2.
32. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
33. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40:W452–7.
34. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* 2009;30:1237–44.
35. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7:e46688.
36. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* 2016;32:2230–2.
37. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013;34:57–65.
38. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun* 2020;11:5918.
39. Koczkowska M, Krawczynska N, Stukan M, Kuzniacka A, Brozek I, Sniadecki M, et al. Spectrum and prevalence of pathogenic variants in ovarian cancer susceptibility genes in a group of 333 patients. *Cancers* 2018;10:442.
40. Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* 2011;108:18032–7.
41. De Brakeleer S, De Greve J, Loris R, Janin N, Lissens W, Sermijn E, et al. Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. *Hum Mutat* 2010;31:E1175–85.
42. Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer* 2015;121:25–33.
43. Ramus SJ, Song H, Dicks E, Tyrer JP, Rosenthal AN, Intermaggio MP, et al. Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. *J Natl Cancer Inst* 2015;107:djv214.
44. Spugnese L, Gabriele M, Scarpitta M, Tancredi M, Maresca L, Gambino G, et al. Germline mutations in DNA repair genes may predict neoadjuvant therapy response in triple negative breast patients. *Genes Chromosomes Cancer* 2016;55:915–24.
45. Caminsky NG, Mucaki EJ, Perri AM, Lu R, Knoll JH, Rogan PK. Prioritizing variants in complete hereditary breast and ovarian cancer genes in patients lacking known BRCA mutations. *Hum Mutat* 2016;37:640–52.
46. Prescott NJ, Lehne B, Stone K, Lee JC, Taylor K, Knight J, et al. Pooled sequencing of 531 genes in inflammatory bowel disease identifies an associated rare variant in BTNL2 and implicates other immune related genes. *PLoS Genet* 2015;11:e1004955.
47. Kaartokallio T, Wang J, Heinonen S, Kajantie E, Kivinen K, Pouta A, et al. Exome sequencing in pooled DNA samples to identify maternal pre-eclampsia risk variants. *Sci Rep* 2016;6:29085.
48. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324:387–9.
49. Bansal V, Gassenhuber J, Phillips T, Oliveira G, Harbaugh R, Villaras N, et al. Spectrum of mutations in monogenic diabetes genes identified from high-throughput DNA sequencing of 6888 individuals. *BMC Med* 2017;15:213.