The contribution of deleterious germline mutations in *BRCA1*, *BRCA2* and the mismatch repair genes to ovarian cancer in the population

Honglin Song1,*, Mine S. Cicek2, Ed Dicks1, Patricia Harrington1, Susan J. Ramus3, Julie M. Cunningham2, Brooke L. Fridley4, Jonathan P. Tyrer1, Jennifer Alsop1, Mercedes Jimenez-Linan5, Simon A. Gayther3, Ellen L. Goode2 and Paul D.P. Pharoah1

1CR-UK Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK, 2Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, 3Department of Preventive Medicine, Keck School of Medicine, USC/Norris Comprehensive Cancer Center, University of Southern California, CA, USA, 4Department of Biostatistics, University of Kansas Medical Center, Kansas, USA and 5Department of Histopathology, Addenbrooke’s Hospital, Cambridge, UK

Received October 31, 2013; Revised April 2, 2014; Accepted April 8, 2014

The aim of this study was to estimate the contribution of deleterious mutations in *BRCA1*, *BRCA2,* MLH1, MSH2, MSH6 and PMS2 to invasive epithelial ovarian cancer (EOC) in the population. The coding sequence and splice site boundaries of all six genes were amplified in germline DNA from 2240 invasive EOC cases and 1535 controls. Barcoded fragment libraries were sequenced using the Illumina GAIIL or HiSeq and sequence data for each subject de-multiplexed prior to interpretation. GATK and Annovar were used for variant detection and annotation. After quality control 2222 cases (99.2%) and 1528 controls (99.5%) were included in the final analysis. We identified 193 EOC cases (8.7%) carrying a deleterious mutation in at least one gene compared with 10 controls (0.65%). Mutations were most frequent in *BRCA1* and *BRCA2*, with 84 EOC cases (3.8%) carrying a *BRCA1* mutation and 94 EOC cases (4.2%) carrying a *BRCA2* mutation. The combined *BRCA1* and *BRCA2* mutation prevalence was 11% in high-grade serous disease. Seventeen EOC cases carried a mutation in a mismatch repair gene, including 10 *MSH6* mutation carriers (0.45%) and 4 *MSH2* mutation carriers (0.18%). At least 1 in 10 women with high-grade serous EOC has a *BRCA1* or *BRCA2* mutation. The development of next generation sequencing technologies enables rapid mutation screening for multiple susceptibility genes at once, suggesting that routine clinical testing of all incidence cases should be considered.

INTRODUCTION

Epithelial ovarian cancer (EOC) is the most fatal gynecological malignancy, resulting in ~140 000 deaths worldwide per year (1). Despite some recent advances in treatment, there have been only slight improvements in survival of patients diagnosed with EOC in over four decades. Approximately 70% of EOC cases are diagnosed with advanced stage disease, in whom 5-year survival is <30%. By contrast, survival is over 90% for patients with Stage I disease (2). Prophylactic salpingo-oophorectomy reduces the risk of ovarian/fallopian tube cancer in carriers of high-penetrance alleles of *BRCA1* or *BRCA2* by 75–96% (3,4). Individual susceptibility to EOC has a substantial inherited genetic component—women with a single-first degree relative with ovarian cancer have a 3-fold greater risk of developing the disease than women without a family history (5). The known ovarian cancer susceptibility genes, which include *BRCA1* and *BRCA2*, are estimated to explain <40% of the excess familial risk of ovarian cancer. Genes other than *BRCA1* and *BRCA2* that confer >20-fold lifetime risk of EOC are unlikely to exist, given that *BRCA1* and *BRCA2* explain most multi-generation, multi-case EOC families (6,7).

Common, low penetrance susceptibility alleles also contribute to familial EOC risk, and genome wide association studies

*To whom correspondence should be addressed. Tel: +44 1223761940; Fax: +44 1223740147; Email: hs310@medschl.cam.ac.uk

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
(GWAS) performed in large-scale, case–control studies with sample sizes in excess of 40,000 subjects have so far identified 11 confirmed ovarian cancer susceptibility alleles, each conferring relative risks of <1.5 (8–14). However, in combination, these alleles account for just 4% of the excess familial ovarian cancer risk (13) suggesting other susceptibility alleles are likely to exist. A wide variety of genetic models have been suggested to account for the ‘missing heritability’ including the possibility that there are several rare alleles that confer relative risks of >3. Evidence for this has emerged through the recent identification of deleterious alleles in RAD51C, RAD51D and BRIP1 that are associated with an increased risk of EOC (15–17). Truncating mutations in the DNA mismatch repair (MMR) genes are also associated with modest risks of EOC (18,19). In particular mutations of MLH1, MSH2, MSH6 and PMS2 are reported to be associated with ovarian cancer as part of the Lynch Syndrome (18,20,21).

Many studies have investigated the contribution of BRCA1 and/or BRCA2 to EOC incidence in the population but most of these have involved fewer than 500 subjects (22). Only two studies have reported BRCA1 and BRCA2 mutation prevalence based on >1000 samples (23,24). Mutation prevalence reported in population-based studies that screened the full coding sequence varies widely from 3.4 to 9.8% for BRCA1 and from 0.6 to 5.7% for BRCA2 (22). A handful of studies have reported the prevalence of deleterious mutations in the MMR genes to be 0.5–3% in EOC (21,25–28). Furthermore, no large studies have evaluated the frequency of deleterious mutations in these genes in non-cancer controls. The aim of this study was to establish the contribution of predicted deleterious mutations in the BRCA1, BRCA2, MLH1, MSH2, MSH6 and PMS2 to ovarian cancer risk in the population using targeted next generation sequencing in two large case–control studies.

RESULTS

Sequence data for BRCA1, BRCA2, MLH1, MSH2, MSH6 and PMS2 were available in 2222 cases and 1528 controls after quality control. The clinical–pathological characteristics of cases in this study are presented in Table 1. Controls were individuals with no known diagnosis of ovarian or any other cancer.

We detected a total of 813 different variants of which 85 (10%) were frameshift indels, 10 (1.2%) were variants predicted by MaxEntScan (29) to affect gene splicing, 37 (4.6%) were nonsense substitutions and 506 (62%) were missense substitutions and 175 (22%) were synonymous substitutions. Of the missense substitutions, 24 had an alternate allele frequency >2% and were not considered deleterious. We classified six rare missense variants in BRCA1 as deleterious, two in BRCA2, one in MLH1 and two in MSH6. Three hundred and forty missense variants were predicted by SIFT/polyphen-2 to be neutral and 131 were predicted to have some functional effect. These were all considered unclassified. One nonsense substitution—BRCA2 K3326X—had a carrier frequency of 2.0% in controls and 2.5% in the cases (P = 0.31), and so was not considered deleterious.

Two hundred and three subjects carried at least one of the 143 different deleterious variants (henceforth mutations) identified. The mutation prevalence was 8.7% in cases (193 out of 2222) and 0.65% in controls (10 out of 1528). Details of each protein truncating mutation and potential deleterious missense/synonymous mutation we identified are given in Supplementary Material, Tables S1 and S2, respectively. In EOC cases, we identified 84 BRCA1 mutation carriers (3.8%), 94 BRCA2 mutation carriers (4.2%), 10 MSH6 mutation carriers (0.45%), 4 MSH2 mutation carriers (0.2%) and 2 MLH1 mutation carriers (0.1%) and 1 PMS2 mutation carrier (0.05%) (Table 2). One case with a mucinous subtype ovarian cancer had a deleterious mutation in both MSH6 and BRCA1; another case with a clear cell subtype ovarian cancer had a deleterious mutation in both BRCA1 and MLH1. Of the deleterious mutations in controls, we identified one deleterious mutation in BRCA1, 4 in BRCA2, 2 in MLH1 and 3 in MSH6 (Table 2).

Thirty-one deleterious mutations were detected in more than one individual, with the most common, 4065_4066delTTGA in BRCA1, being found in seven individuals. The BRCA1 mutation 5266dupC (known as 5382insC), which is relatively common in individuals of Eastern European origin, was identified in five individuals. Of the total 195 deleterious mutations in cases, 127 (65%) were insertion or deletion frameshifts, 41 (21%) were nonsense substitutions, 12 (6.2%) were variants situated near exon/intron boundary that were predicted to affect gene splicing, and 15 (7.7%) were previously reported pathogenic/likely pathogenic missense substitutions (30–33) (Supplementary Material, Table S2). In controls, four mutations were frameshift indels, four were nonsense mutations and two variants that were predicted to affect splicing.

Previous studies have indicated that mutations located in specific regions of BRCA1 (cDNA 2282–4066) (7, 34) and BRCA2 (cDNA 2807–6401, termed the Ovarian Cancer Cluster Region or OCCR (7,35) are associated with a relatively higher risk of ovarian cancer compared with breast cancer. In our study, the proportion of BRCA1 mutations located in this region was similar to that would be expected by chance if mutations occurred evenly across the coding region (33 versus 32%). The proportion of mutations in the OCCR region of BRCA2 was significantly higher than that expected by chance (54 versus 35%, P = 0.0017).

Table 1. Characteristics of the ovarian cancer subjects (N = 2222)

<table>
<thead>
<tr>
<th>SEARCH (N = 1310)</th>
<th>Mayo Clinic (N = 912)</th>
<th>Total (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age at diagnosis, years (range)</td>
<td>55.9 (19–74)</td>
<td>62.5 (23–91)</td>
</tr>
<tr>
<td>Morphology (n, %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serous</td>
<td>600 (46)</td>
<td>677 (74)</td>
</tr>
<tr>
<td>Endometrioid</td>
<td>212 (16)</td>
<td>110 (12)</td>
</tr>
<tr>
<td>Clear cell</td>
<td>137 (10)</td>
<td>55 (6.0)</td>
</tr>
<tr>
<td>Mucinous</td>
<td>132 (10)</td>
<td>25 (3.0)</td>
</tr>
<tr>
<td>Other/unknown</td>
<td>229 (16)</td>
<td>45 (5.0)</td>
</tr>
<tr>
<td>Stage (n, %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>431 (33)</td>
<td>141 (15)</td>
</tr>
<tr>
<td>II</td>
<td>121 (9)</td>
<td>51 (5.6)</td>
</tr>
<tr>
<td>III</td>
<td>368 (28)</td>
<td>538 (59)</td>
</tr>
<tr>
<td>IV</td>
<td>66 (5.0)</td>
<td>171 (19)</td>
</tr>
<tr>
<td>Unknown</td>
<td>324 (25)</td>
<td>11 (1.2)</td>
</tr>
<tr>
<td>Grade (n, %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low grade</td>
<td>169 (13)</td>
<td>128 (14)</td>
</tr>
<tr>
<td>High grade</td>
<td>714 (55)</td>
<td>754 (83)</td>
</tr>
<tr>
<td>Unknown</td>
<td>427 (33)</td>
<td>30 (3)</td>
</tr>
</tbody>
</table>
entiated tumours, four were high-grade endometrioid tumours, two MMR gene mutations were identified. Five EOC cases with 11% compared with 5% for other subtypes. A total of 17 BRCA2 mutation prevalence in high-grade serous disease was reported a family history of ovarian cancer only and 18 reported history of breast cancer only was reported by 302 cases, 88 low-grade-endometrioid and high-grade-mucinous tumour. were high-grade serous tumours and one each of low-grade serous, of good prognosis tumours (non-serous) will be higher. In total, second, SEARCH included prevalent cases and so the proportion had central pathology review and only 5% were unspecified; SEARCH cases had 16% unspecified subtype derived from characteristics of ovarian cancer

We compared the frequency of deleterious mutation carriers in cases and controls using an odds ratio. The odds ratio was 60 (95% confidence interval 10–2100) for BRCA1, 17 (6.3–63) for BRCA2 and 2.3 (0.83–8.2) for MMR gene mutation carriers. The average cumulative risks of ovarian cancer by age 80 years are estimated to be 61% (15–99%) for BRCA1, 24% (10–62%) for BRCA2 and 3.7% (1.4–13%) for MMR genes.

**Associations with clinical and histopathological characteristics of ovarian cancer**

**BRCA1** mutation carriers were diagnosed at a median age of 52 years (range 33–82 years); this compared with 57 years (range 33–84 years) for BRCA2 carriers, 54 years (range 43–73 years) for MMR gene mutation carriers and 59 years (range 19–91) for non-mutation carriers (Table 3). The proportion of cases of the serous subtype was higher in the Mayo Clinic case-series (74%) than the SEARCH case-series (46%) (Table 1). Of these, 34% of SEARCH cases (N = 451) and 72% of Mayo Clinic cases (N = 654) were classified as high-grade serous subtype. There are two primary reasons for this: first, the SEARCH cases had 16% unspecified subtype derived from routine pathology report, whereas the Mayo Clinic cases all had central pathology review and only 5% were unspecified; second, SEARCH included prevalent cases and so the proportion of good prognosis tumours (non-serous) will be higher. In total, 57% of the cases were diagnosed with serous tumours; the remainder was a mix of endometrioid, clear cell, mucinous or other/unknown subtypes (Table 1). The prevalence of deleterious mutations varied by histopathological subtype and by study (Table 2). In the combined results, 122 of the 178 BRCA1 and BRCA2 mutation carriers (69%) were identified in patients diagnosed high-grade serous EOC, and BRCA1/BRCA2 mutation prevalence in high-grade serous disease was 11% compared with 5% for other subtypes. A total of 17 MMR gene mutations were identified. Five EOC cases with MMR gene mutations were clear cell tumours, three were undifferentiated tumours, five were high-grade endometrioid tumours, two were high-grade serous tumours and one each of low-grade serous, low-grade-endometrioid and high-grade-mucinous tumour.

Information on breast and/or ovarian cancer history in first-degree relatives was available for 1862 of 2222 cases. A family history of breast cancer only was reported by 302 cases, 88 reported a family history of ovarian cancer only and 18 reported a family history of both breast and ovarian cancer. Of the 408 cases reporting a family history, 78 (19%) carried a deleterious variant in either BRCA1 or BRCA2 compared with 70 mutation carriers out of 1454 cases (4.8%) with no family history (Table 4). Information on family history was available for 11 of the 17 MMR gene mutation carriers; five of these women reported a first- or second-degree relative with endometrial cancer and three reported a first- or second-degree relative with colorectal cancer.

**Association between rare missense variants and ovarian cancer risk**

In addition to the probable deleterious variants, we identified 471 unclassified missense variants across all six genes of which 131 are predicted to have a functional effect based on SIFT and polyphen-2. We compared the relative burden of the possibly functional variants in cases and controls for each gene using the rare admixture maximum likelihood test (RAML) (36). We found little evidence for association of rare missense variation in any of these genes and ovarian cancer risk (all $P > 0.05$).

**DISCUSSION**

This is the largest population-based ovarian cancer study to estimate the prevalence of mutations in BRCA1, BRCA2 and the MMR genes reported to date. Overall, 8.7% of epithelial ovarian cancer cases had a mutation in one of the six genes.

---

**Table 2.** Number of mutations identified in six genes by histology subtype

<table>
<thead>
<tr>
<th></th>
<th>Control (n = 1528)</th>
<th>SEARCH cases (n = 1310)</th>
<th>Mayo Clinic cases (n = 912)</th>
<th>Cases combined (n = 2222)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>High-grade</td>
<td>Other</td>
<td>High-grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>serous</td>
<td></td>
<td>serous</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>1518</td>
<td>388</td>
<td>86</td>
<td>803</td>
</tr>
<tr>
<td>Mutation carrier</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>10</td>
<td>0.6</td>
<td>63</td>
<td>14</td>
</tr>
<tr>
<td>BRCA1</td>
<td>1</td>
<td>0.07</td>
<td>25</td>
<td>5.5</td>
</tr>
<tr>
<td>BRCA2</td>
<td>4</td>
<td>0.3</td>
<td>38</td>
<td>8.4</td>
</tr>
<tr>
<td>MMR*</td>
<td>5</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>1528</td>
<td>100</td>
<td>451</td>
<td>100</td>
</tr>
</tbody>
</table>

*a One case had both BRCA1 and MLH1 mutation and another case had both BRCA1 and MSH6 mutation.

**Table 3.** Mutation status by age of disease onset

<table>
<thead>
<tr>
<th>Mutation status</th>
<th>Age at diagnosis (n, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;40 years</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>94 (5)</td>
</tr>
<tr>
<td>Mutation carrier</td>
<td>6 (3)</td>
</tr>
<tr>
<td>BRCA1</td>
<td>4 (5)</td>
</tr>
<tr>
<td>BRCA2</td>
<td>2 (2)</td>
</tr>
<tr>
<td>MLH1/MSH2/MSH6/ PMS2*</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

*a One case had both BRCA1 and MLH1 mutation and another case had both BRCA1 and MSH6 mutation.
compared with 0.65% in controls. We have almost certainly underestimated the true mutation prevalence in this series of genes, as some deleterious mutations will have been missed using our next generation sequencing approach. For example, we did not sequence the entire coding sequence in all individuals; the mean coverage for the targeted gene regions was 84% ranging from 71 to 92% for the six genes at read depth at least 15 (Supplementary Material, Table 3). In addition, some missense substitutions of uncertain significance, which were excluded from the prevalence estimates, may be true deleterious mutations. Finally, our PCR based method would not have been able to detect large genomic rearrangements, which we have previously shown account for ~10% of mutations in BRCA1/BRCA2 (7) and 5–20% of all MMR gene mutations in Lynch syndrome families (37,38). By extrapolation, we calculate that this study may have underestimated the true prevalence of deleterious BRCA1, BRCA2 and MSH6 mutations by about 17, 24 and 30% respectively which would increase the deleterious mutation prevalence estimates in EOC to 4.6% for BRCA1, 5.6% for BRCA2 and 0.6% for MSH6.

BRCA1 or BRCA2 mutations were more common in the high-grade serous ovarian cancer subtype. This is consistent with the previous findings that DNA double strand break repair is associated with high-grade serous EOC (39). The estimated true prevalence of BRCA1 and BRCA2 in high-grade serous ovarian cancer is 14.1%: this is consistent with The Cancer Genome Atlas (TCGA) project, which identified germline BRCA1/BRCA2 in 16% of high-grade serous EOC cases (40).

The frequency of mutations in MSH6 for EOC cases (10/2222 or 0.45%) is similar to that reported by Walsh et al. (2/360 or 0.55%) (28). Similar to this study, and as previously reported for epithelial ovarian cancer cases in Lynch syndrome families (20), mutations in MMR genes were mostly of the non-serous subtype.

The frequency of mutations in BRCA1 and BRCA2 is somewhat lower than that reported in the two previously largest published studies (8% in SEARCH, 8% in Mayo Clinic cases, 13% in Zhang et al. (23) and 14% in Alsop et al. (24)). However, the published data include data for large genomic rearrangements. In addition the ancestry of the populations is different – the three variants common in the Ashkenazi Jewish population accounted for 11–14% of all deleterious variants identified in the two large published case series compared with 3% in our study (Supplementary Material, Table 4). The deleterious variant prevalence in the high-grade serous subtype was substantially higher in the Australian cases series (23%) than in SEARCH (14%), Mayo Clinic cases (9%) or the Canadian case series (18%). The clinical characteristics of the four case series are broadly similar apart from a smaller proportion of serous cases in SEARCH and a higher proportion of cases diagnosed at age 60 or older in Mayo Clinic cases (Supplementary Material, Table 5). Thus, reasons for this difference are not clear.

In summary, we estimate that BRCA1 and BRCA2 together are major contributors to at least one in every 12 invasive EOC cases, and one in every 10 high-grade serous EOC cases. Mutations in MMR genes are responsible for at least one in every 131 EOC cases. The recently published NICE Guideline for the management of women with a family history of breast cancer has recommended that individuals with a 10% or greater probability of carrying a deleterious mutation in BRCA1 or BRCA2 should be offered genetic testing (NICE clinical guidelines Issued: June 2013, CG164). Thus, all women with high-grade serous ovarian would be eligible for BRCA1 and BRCA2 testing under the NICE Guideline. Given that PARP inhibitors have been shown to improve progression-free survival among patient with platinum-sensitive, relapsed, high-grade serous ovarian cancer (41), such testing may have important implications for the patient. The potential implications for the family of BRCA1 or BRCA2 mutation carriers are clearer, as cascade testing of unaffected female relatives will identify unaffected carriers for whom the benefits of prophylactic salpingo-oophorectomy in reducing risk of incident ovarian, fallopian tube and peritoneal cancer are well established. Furthermore, it will be feasible to use a targeted sequencing approach for clinical testing of multiple genes in a single assay to replace the conventional clinical genetic testing which was done gene by gene with relatively high cost. The widespread application of panel sequencing in unselected women in the UK population and the provision of prophylactic salpingo-oophorectomy to mutation carriers have the potential to reduce the incidence of epithelial ovarian cancer by 8%.

## MATERIALS AND METHODS

### Study subjects

Confirmed invasive EOC cases and unaffected controls were from two case-controls studies: the population-based SEARCH study...
(1321 cases/1389 controls) from the United Kingdom, and the hospital-based Mayo clinic study (919 cases/146 controls) from the USA. The studies have been previously described (42). Briefly, the SEARCH ovarian cancer study comprises 1321 invasive epithelial ovarian cancer cases <75 years from East Anglian West Midlands and Trent regions of England. Prevalent cases diagnosed between 1991–1998; incident cases diagnosed 1998 onwards. All eligible cases were identified through Cancer Registry. Pathology information was abstracted from routine pathology reports. Cancer free controls are from a population-based controls series from East Anglia (UK) and frequency matched to cases on age. The Mayo Clinic ovarian cancer study comprises 919 cases diagnosed from 2000 onwards at Mayo Clinic, for whom a diagnosis of histologically confirmed primary epithelial ovarian cancer was ascertained within one year of consent. The pathology of all eligible cases were centrally reviewed by a group of gynecologic pathologists. Controls were frequency matched to cases on race and age from the same study area. Based on data from single nucleotide polymorphism genotyping, 2214 cases and 1518 controls were of European ancestry (13). For quality control purposes we also included 43 duplicate samples and 34 positive controls known to be BRCA1 or BRCA2 mutation carriers. Both studies have ethics committee approval, and all study subjects provided written, informed consent.

Sequencing library preparation and sequencing
We used the 48.48 Fluidigm Access Arrays™ for target sequence enrichment. This method uses a four-primer chemistry that allows the addition of barcode and adapter sequences to small target regions (<210 bp) during up to 10-plex PCR amplification. A total of 338 primer-pairs were designed to cover the exons and splice sites of BRCA1, BRCA2, MLH1, MSH2, MSH6 and PMS2 (Supplementary Material, Table S3). The combined sequencing target for these six genes was about 28 kb with some of the sequence being covered by multiple amplicons. The primer design achieved more than 94% coverage of the coding exons and potential splice site (i.e. 20 bp in the intron for 3' acceptor sites and 6 bp in the intron for donor 5' sites) sequence for each gene. Sequencing library preparation used 1.25 μl germline DNA at 75 ng/μl according to the manufacturer’s protocol (Fluidigm, San Francisco, CA, US). Sequencing libraries were quantified using a KAPA library quantification kits with specific probes for the ends of the adapters according to the manufacturer’s protocol (KapaBiosystems, Boston, MA, US). The sequence library was sequenced using either the single end sequencing on the Illumina GAII (1443 samples) or paired end sequencing on the Illumina HiScan (2332 samples) according to the manufacturer’s protocol (Illumina Inc., San Diego, CA, US). Sample barcoding enabled 384 individual’s sample to be sequenced on each lane of the sequencer.

Data analysis
Sequenced reads were de-multiplexed using standard Illumina software. We used the Burrows-Wheeler Aligner (BWA) (43) for sequencing read alignment against the human genome reference sequence (GRCh37, UCSC hg19). The Genome Analysis Toolkit (GATK) (44) was then used for base quality-score recalibration, local insertion/deletion (indel) realignment, substitution and indel discovery. Variants were only considered if they satisfied the set of recommended GATK filters, as described in the GATK best practices guide. ANNOVAR (45) was used to annotate the sequence variation detected. We used PolyPhen-2 (46) and SIFT (47) to predict the function of missense variants. We used MaxEntScan (29) to predict the pathogenic potential of putative splicing mutations in sequences from 3 bp in the exon to 20 bp in the intron for 3' acceptor sites and 3 bp in the exon and 6 bp in the intron for the donor 5' sites. MaxEntScan provides a score for the strength of the splice site and enables the scores for the consensus and variant sequences to be compared. It has been suggested that a score for a variant sequence 20% lower than the consensus sequence score as likely to result in abnormal splicing (48). However, after applying this criterion we classified multiple splice site variants as deleterious in both cases and controls. We therefore applied a more stringent criterion of 40% lower, which identified 17 variants, 16 of these occurring in cases.

Variant alternate allele frequency was defined as the fraction of alternative allele reads compared with the total number of reads at the variant locus. We used the variants called in the sequence data from the 43 duplicate samples and additionally the BRCA1 and BRCA2 positive controls to define thresholds, based on coverage and alternate allele frequency, for definitive variant calling in the full data set. Variants with depth <15 were not called. Alternate allele heterozygotes were called if (i) depth ≥500 and alternative allele frequency ≥10%; (ii) 250 ≤ depth < 500 and alternative allele frequency ≥15%; (iii) 30 ≤ depth < 250 and alternative allele frequency ≥20%; (iv) 15 ≤ depth < 30 and alternative allele frequency ≥30%. Samples with fewer than 80% of amplicons covered at a read depth of ≥15 were excluded from subsequent analyses (a total of 7 controls and 18 cases failed in both sequencing runs). We defined deleterious variants as those that are predicted to result in protein truncation (frame shift indels, consensus splice site substitutions and nonsense substitutions) or those missense mutations that have been previously reported or classified (e.g. http://www.insight-group.org/ for MMR missense variants classification) or predicated by MaxEntScan as deleterious.

The median read depth for our targeted sequencing was 102 (IQR 91–115) and 548 (IQR 492–626) for Illumina GAII and HiSeq, respectively. Eighty-two percent of targeted sequence base had read depth ≥30 and 85% had read depth ≥15. The average sequencing coverage for these six genes is summarized in Supplementary Material, Table S3. The concordance rate for the 43 duplicates was 100%. Thirty-one of 34 (91%) known BRCA1 or BRCA2 positive controls were called correctly. Three mutations were missed because the read depth of the relevant amplicon was <15.

Mutation validation
Two hundred and five potentially deleterious variants were identified. We inspected the sequence alignments for all the deleterious variants using the Integrative Genomic viewer (IGV) (49). Ninety-four percent (193/205) of these had an alternative allele read frequency ≥30%. We validated 24 deleterious variants, including all variants with an alternative allele read frequency of <30% (n = 12) by PCR amplification and Sanger sequencing.
Unclassified variant analysis

We also identified multiple unclassified, but putative functional missense variants with allele frequencies of <2% in the samples. Because the statistical power to detect single rare alleles by association, even with large sample sizes is modest, we performed burden tests, which combines the information across multiple variants to increase statistical power. We classified variants with frequency ≤2% into three groups: (i) deleterious variants as defined previously; (ii) variants predicted to have a damaging effect on protein function—SIFT score <0.05 and polyphen-2 classified as possibly damaging/probably damaging; (iii) variants with probable benign effects. We used the RAML (36) to test for the association of uncommon missense variants (MAF ≤ 2%) with ovarian cancer risks on a gene-by-gene basis. RAML takes account of variants that increase or decrease risk. Only subjects with a call rate >80% and variants with a call rate >80% with genotype frequencies consistent with Hardy–Weinberg equilibrium (P > 10⁻⁵) were included in these analyses.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank all the study participants who contributed to this study and all the researchers, clinicians and technical and administrative staff who have made possible this work. In particular, we thank Marie Mack, Craig Luccarini, Caroline Baynes, the SEARCH team and Eastern Cancer Registration and Information Centre (SEARCH).

Conflict of Interest statement. None declared.

FUNDING

This work was funded by Cancer Research UK (C490/A10119 and C490/A10124) (SEARCH) and National Institute of Health (R01 CA122443), P50-CA136393, Fred C. and Katherine B. Andersen Foundation (Mayo Clinic study), Cancer Research UK grant C1005/A12677 and National Institute of Health grant R01 CA122443, P50-CA136393, Fred C. and Katherine B. Andersen Foundation (Mayo Clinic study), Cancer Research UK grant C1005/A12677 and National Institute of Health grant R01 CA178535 (USC), and the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge.

REFERENCES


