A targeted genetic association study of epithelial ovarian cancer susceptibility

Permalink
https://escholarship.org/uc/item/9ww2j29f

Journal
ONCOTARGET, 7(7)

ISSN
1949-2553

Authors
Earp, M
Winham, SJ
Larson, N
et al.

Publication Date
2016-02-16

License
CC BY 4.0

Peer reviewed
A targeted genetic association study of epithelial ovarian cancer susceptibility

Madalene Earp¹, Stacey J. Winham², Nicholas Larson², Jennifer B. Permuth³, Hugues Sicotte², Jeremy Chien⁴, Hoda Anton-Culver⁵, Elisa V. Bandera⁶, Andrew Berchuck⁷, Linda S. Cook⁸, Daniel Cramer⁹,¹⁰, Jennifer A. Doherty¹¹, Marc T. Goodman¹², Douglas A. Levine¹³, Alvaro N.A. Monteiro³, Roberta B. Ness¹⁴, Celeste L. Pearce¹⁵, Mary Anne Rossing¹⁶,¹⁷, Shelley S. Tworoger¹⁰,¹⁸, Nicolas Wentzensen¹⁹, Maria Bisogna¹³, Louise Brinton¹⁹, Angela Brooks-Wilson²⁰,²¹, Michael E. Carney²², Julie M. Cunningham²³, Robert P. Edwards²⁴, Zachary C. Fogarty², Edwin S. Iversen²⁵, Peter Kraft²⁶, Melissa C. Larson², Nhu D. Le²⁷, Hui-Yi Lin³, Jolanta Lissowska²⁸, Francesmary Modugno²⁴,²⁹,³⁰, Kirsten B. Moysich³¹, Sara H. Olson³², Malcolm C. Pike¹⁵,³¹, Elizabeth M. Poole¹⁸, David N. Rider², Kathryn L. Terry⁹,¹⁰, Pamela J. Thompson¹², David van den Berg¹⁵, Robert A. Vierkant², Allison F. Vitonis⁹, Lynne R. Wilkens³³, Anna H. Wu¹⁵, Hannah P. Yang¹⁹, Argyrios Ziogas³⁴, Catherine M. Phelan³, Joellen M. Schildkraut³⁵,³⁶, Yian Ann Chen³, Thomas A. Sellers³, Brooke L. Fridley³⁷ and Ellen L. Goode¹

¹ Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA
² Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA
³ Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
⁴ Department of Cancer Biology, University of Kansas Cancer Center, Kansas City, KS, USA
⁵ Department of Epidemiology, University of California Irvine, Irvine, CA, USA
⁶ Rutgers Cancer Institute of New Jersey and Robert Wood Johnson Medical School, New Brunswick, NJ, USA
⁷ Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA
⁸ Division of Epidemiology and Biostatistics, University of New Mexico, Albuquerque, NM, USA
⁹ Obstetrics and Gynecology Epidemiology Center, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA
¹⁰ Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA
¹¹ Section of Biostatistics and Epidemiology, The Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
¹² Samuel Oschin Comprehensive Cancer Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA
¹³ Gynecology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
¹⁴ The University of Texas School of Public Health, Houston, TX, USA
¹⁵ Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
¹⁶ Department of Epidemiology, University of Washington, Seattle, WA, USA
¹⁷ Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
¹⁸ Channing Division of Network Medicine, Harvard Medical School and Brigham and Women’s Hospital, Boston, MA, USA
¹⁹ Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
²⁰ Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada
²¹ Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada
²² Clinical and Translational Research Program, University of Hawaii Cancer Center, Honolulu, HI, USA
²³ Department of Laboratory Medicine and Pathology, Division of Experimental Pathology, Mayo Clinic, Rochester, MN, USA
²⁴ Department of Obstetrics, Gynecology and Reproductive Sciences, Division of Gynecologic Oncology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
²⁵ Department of Statistical Science, Duke University, Durham, NC, USA
²⁶ Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA, USA
INTRODUCTION

Epithelial ovarian cancer (EOC) is the second most common gynecologic cancer in the US, but it leads in deaths owing to its tendency to be diagnosed in the late stages of disease [1]. EOC is composed of five major histologic types [2]: high-grade serous carcinoma (HGSC), accounting for most cases (~70%); and the rarer clear cell, endometrioid, mucinous, and low-grade serous carcinomas (LGSC). Known rare mutations in DNA repair and mismatch repair genes are thought to account for 10%-15% of all EOCs [3-9]. Common alleles identified by genome-wide association studies (GWAS) are thought to account for an additional 3%-4% of EOC risk [10-17]. Still, much of about the heritability of EOC remains unaccounted for. Here, we sought to identify additional EOC susceptibility variants through direct genotyping and analysis of EOC cases and controls from 13 independent studies. We targeted variants based on innovative pilot studies, hypothesizing that previously ungenotyped variants may be responsible for a proportion of the unexplained EOC susceptibility.
RESULTS

Known EOC susceptibility regions

One goal of this project was to compare the relative strength of the associations between known and novel variants within the first eleven published EOC risk loci[10-12, 14, 18] (Supplemental Table 1). The variants most strongly associated with EOC risk in this study (all histology or HGSC) are given in Table 1 and plotted regionally in Figure 1, Figure 2, and Supplemental Figure 1. Compared to published variants, novel variants were more strongly associated with susceptibility of all histologies of EOC at nine loci (2q31, 3q25, 5p15, 8q21, 8q24, 10p12, 17q12, 17q21.31, and 19p13) (Table 1); all but three (8q21, 17q21.31, and 19p13) are in moderate LD (r^2 > 0.4) with known variants (Supplemental Figure 1). At the 3q25 locus variant rs62273902 (p_{all-histology} = 2 x 10^{-8}) was the most strongly associated variant (Figure 1), and at the 17q21.31 locus variant rs2532240 (p_{all-histology} = 3 x 10^{-7}) was the most strongly associated variant (Figure 2). In the HGSC only analysis, novel variants were more strongly associated with susceptibility at seven loci (2q31, 3q25, 8q24, 10p12, 17q12, 17q21.31, and 19p13; Table 1); all but two (2q31, 17q21.31) are in moderate LD (r^2 > 0.4) with known variants (Supplemental Figure 1). With two exceptions (noted below), novel variants were common (minor allele frequency (MAF)>0.05), in the intron of genes or intergenic, in moderate-to-strong LD with known variants, and conferred modest effects on susceptibility. One exception was the association of rare intergenic variant rs74955251 at 8q21 (MAF_{overall} = 0.00028, OR = 3.9 x 10^{-4}, 95% confidence interval [CI]: 3.4 x 10^{-4} to 4.6 x 10^{-4}, p_{all-histology} = 4 x 10^{-4}). Given its rarity, rs74955251 requires assessment in a much larger sample of cases and controls. A second exception was the association of common (MAF_{overall} = 0.50) missense variant rs23629356 in the gene ANKLE1 at 19p13 (OR = 0.91, 95% CI: 0.87-0.97, p_{all-histology} = 2 x 10^{-3}, protein change Leu184Trp).

Beyond known EOC susceptibility regions

We targeted 5,320 variants which showed suggestive association with susceptibility in a pilot-scale whole genome sequence analysis that compared germline sequence of EOC patients (N= 19) to 1000 GP participants (N= 174). No novel variants reached genome-wide significance for association with EOC risk overall or HGSC (p ≤ 5x10^{-8}), nor were significant after Bonferroni correction (p ≤ 9x10^{-4}). Nonetheless, the risk estimates generally were in the expected direction based on pilot data, and several variants merit investigation in larger case-control collections (Table 2). For example, among variants targeted because they were present only in EOC germline sequence data (WGS EOC+ in Table 2), the most strongly associated risk variant was rs138643956 (OR= 3.68; p_{HGSC} = 2 x 10^{-4}). For variants selected because they were absent from whole genome sequenced EOC cases (WGS EOC- in Table 2), the most associated variant was rs9380516 (OR= 0.83; p_{HGSC} = 6 x 10^{-8}); in the current study this variant showed a case MAF of 0.15, suggesting that this was a missed variant in the pilot sequencing study. For variants targeted which were present in whole genome sequenced EOC cases and in 1000 GP data, but differed in HGSC (WGS EOC↑ and WGS EOC↓ in Table 2), the current analyses were generally consistent, including rs117841616 on chromosome 20 (p_{all-histology} = 2 x 10^{-4}) and rs240783 (p_{HGSC} = 8 x 10^{-4}) on chromosome 6. In general, very few variants targeted based on suggestive association in pilot sequence study had appreciable MAF differences (case vs. control) in the current genotyping study. As expected due to small sample size, we observed that case MAF estimates in our pilot whole genome sequencing study were both inflated and deflated compared to the current study.

Finally, among NF-κB-related variants and those hypothesized to associate with endometrioid EOC risk, the most suggestive results for variants which disrupt NF-κB binding [19, 20] were intergenic variants rs10143322 on chromosome 14 (p_{all-histology} = 3 x 10^{-5}) and rs6092485 on chromosome 20 (p_{HGSC} = 7 x 10^{-4}) (Table 2). If Bonferroni correction for the number of NF-κB binding site variants tested is applied, the threshold for statistical significance is p < 4 x 10^{-5} (p = 0.05/1,302), and this single variant, rs10143322, is declared significant; using experiment-wide and certainly genome-wide multiple testing corrections, it is not significant. Among variants previously identified in a pilot GWAS of endometrioid EOC, the most significant variants were intronic variant rs2638653 (p_{endometroid} = 1 x 10^{-4}) on chromosome 8 (Table 2), and intergenic rs9264042 on chromosome 6 (p_{all-histology} = 5 x 10^{-4}). These modest associations may also warrant follow-up in larger studies.

DISCUSSION

The objective of this study was to test whether novel variants identified through a combination of approaches were associated with EOC susceptibility. We took an innovative approach to the selection of variants, including the use of whole genome sequencing data to target novel variants correlated with known GWAS risk variants, comparison of sequencing EOC cases to 1000 GP participants beyond these regions, NF-κB functional data, and GWAS analysis of EOC cases with endometrioid histology.

In nine of the eleven susceptibility regions investigated, novel variants were more highly associated with all histology EOC risk than previously reported variants, and, in the HGSC only analysis, novel variants...
were more strongly associated with susceptibility at seven loci. Further work on these variants may provide more biological insight. For example, at the 3q25 locus, the novel variant rs62273902 (all histology) coincides with a genomic sequence that appears functionally active in a range of cell lines and tissues relevant to EOC, including ovary, as assayed by the Roadmap Epigenomics Mapping Consortium (REMC), http://www.epigenomeweb.org/. rs62273902 resides within a DNase peak, an active transcription start site (TSS), and multiple proteins across diverse tissues bind the sequence spanning this variant. rs62273902 is therefore a good functional candidate variant at this locus. As well, at 17q21.31, the novel variants (rs2532240 in all histologies and rs3785880 in HGSC-only analysis) are separated by 272 kb and not correlated with each other or the previously reported variant rs1294266. rs2532240 coincides with a chromatin region marked as a weak/poised enhancer in several tissues, including ovary (REMC[21]); however, it does not overlap transcription factor binding sites (TFBS) or DNase peaks. rs1294266 does not coincide with promoter or enhancer regions in tissues relevant to EOC in the REMC data, suggesting it is unlikely to be functionally relevant. The 17q21.31 variants are located in a large region of strong LD previously identified as the “17q21.31 inversion” (~900kb long), which exists either as a direct (H1) or inverted (H2) haplotype in the European population [14, 15, 22]. Further investigation of how these variants might impact EOC risk is needed.

Of critical note, a large EOC meta-GWAS with imputation to revised Phase I 1000 GP data was recently completed with over 23,000 cases and 35,000 controls of the Ovarian Cancer Association Consortium, including many of the participants in the current analysis. We inspected our most associated variants from the 11 known susceptibility regions in an online look-up of results based on these data (http://apps.ccge.medschl.cam.ac.uk/ consortia/ocac/contact/contact.html). In general, the variants reported here were highly ranked in the EOC meta-GWAS data (i.e., in the top 50 most associated variants in the regions we defined). At 8q24, the novel directly genotyped variant presented here (rs1400482) was the most associated variant in the larger imputation-based study. At 3q25, 10p12, 17q21.31, and 17q21.32,

![Figure 1: Novel variant rs62273902 in the 5'-untranslated region of LEK1 has the strongest association signal at 3q25. Regional association plot for variants genotyped at 3q25 in all EOC histologies cases \((N = 4,973)\) and controls \((N = 5,640)\). Linkage disequilibrium between rs62273902 and each variant is estimated using data from 5,640 controls and indicated by the color scheme. The previously reported risk variant rs2665390 in this region (Goode, 2010 #23) was not genotyped; rs344008 \((p = 5 \times 10^{-6})\) is indicated in its place to allow comparison of the novel (rs62273902) and known (rs2665390) most associated variants \((r^2 = 1\) for rs344008 and rs2665390 in 1000 Genomes Project phase 1 European data).
the most significant variant in the current study was not among the most significant in the imputation-based study. Nonetheless, novel variants at 3q25 and 17q21.32 remained highly significant (rs62273902 at 3q25 $p_{\text{meta}} = 2 \times 10^{-28}$, and rs9303542 at 17q21.32 $p_{\text{meta}} = 3 \times 10^{-12}$). Although novel variants at 3q25 and 17q21.32 remained highly significant (rs62273902 at 3q25 $p_{\text{meta}} = 2 \times 10^{-28}$, and rs9303542 at 17q21.32 $p_{\text{meta}} = 3 \times 10^{-12}$), at these two regions as well as at 10p12 and 17q21.31, the imputation-based study revealed stronger associations with other variants.

Among variants genotyped based on our pilot study comparing whole genomes of EOC cases and 1000 GP participants, none were significant after multiple testing correction for 5,320 variants. Despite our sample size (4,973 cases and 5,640 controls), power to detect associations with low MAF variants was limited. Variants in NF-κB binding sites were also not associated with EOC risk at genome-wide significance. Noting the debate regarding the use of $p=5 \times 10^{-8}$ as the threshold for statistical significance when evaluating potentially functional variants with presumed higher prior probability[23], Bonferroni correction for the number of NF-κB binding site variants yields one statistically significant variant (rs10143322, $p = 3 \times 10^{-5}$). rs2638653, a variant selected based on an unpublished GWAS of endometrioid EOC, and found here to be the variant most associated with endometrioid EOC risk ($p = 1 \times 10^{-4}$), coincides with chromatin marked as being an active promoter in ovary tissues (of PSD3), but not an enhancer or DNase site. Interestingly, loss of heterozygosity on chromosome 8p22, where this variant is located, is common in EOC tumors, and reduced expression of genes in this region has been found to negatively impact survival in EOC [24].

In summary, we developed a diverse panel of previously ungenotyped variants to directly test for association with EOC susceptibility in 4,973 EOC cases and 5,640 controls from 13 independent studies. Our innovative approach to variant selection included the first use of whole-genome sequencing data from EOC cases in novel variant discovery. At several EOC

**Figure 2:** Novel variant rs2532240 has the strongest association signal at 17q21.31. Regional association plot for variants genotyped at 17q21.31 in all EOC histologies cases ($N = 4,973$) and controls ($N = 5,640$). The most associated variant was rs2532240 ($p = 3 \times 10^{-7}$). Linkage disequilibrium between rs2532240 and each variant is estimated using data from 5,640 controls and indicated by the color scheme. The previously reported risk variant rs12942666 in this region [Permuth-Wey, 2013 #28] was not genotyped, but rs117169618 ($p = 5 \times 10^{-4}$) is indicated in its place to allow comparison of the novel (rs2532240) and known (rs12942666) most variants ($r^2 = 0.8$ for rs117169618 and rs12942666 in 1000 Genomes Project phase 1 European data).
susceptibility regions, we report novel risk variants for further association and functional investigation. Beyond known regions, this first pass at using whole genome sequencing pilot analyses, although underpowered, also yielded variants of potential interest (rs138643956 and rs117841616). The key strength of this report is the use of direct genotyping of novel variants, some rare, while its key limitation is an inability to more comprehensively examine rare variation. Larger scale genotyping and/or improved genotype imputation accuracy will facilitate further scrutiny of the variants highlighted here.

### MATERIALS AND METHODS

#### Study participants

Study participants were drawn from 13 independent EOC case-control studies of the Ovarian Cancer Association Consortium and were restricted to women of European ancestry. Characteristics of the contributing studies are given in Supplemental Table 2 and have been described previously [18]. Cases (N=4,973) consisted of women aged 18 and older with a pathologically confirmed primary invasive EOC, fallopian tube cancer, or primary peritoneal cancer; controls (N=5,640) were matched by age, site, and three European principal components.
Genotyping array

A total of 17,439 germline DNA variants were genotyped using a customized Affymetrix Axiom Exome array (Affymetrix Corporation, Santa Clara, CA). These variants were drawn from four discovery categories: 1) from eleven known EOC susceptibility regions (N = 6,948; Supplemental Table 3) [10-14, 18], identified by in silico fine-mapping and a small germline whole genome sequencing study of EOC cases, 2) variants outside these eleven regions which showed suggestive association in pilot whole genome sequencing of serous EOC cases (compared to 1000 Genomes Project [GP] data) (N = 7,189), 3) variants with a hypothesized role in disrupted binding of NF-κB transcription factors, which are known to have central roles in immune and inflammatory responses and cancer [19, 20, 25, 26] (N = 1,302), and 4) the top associated variants from a pilot GWAS of endometrioid EOC (N = 2,000). See the Supplemental Methods for more detail on the selection of these variants.

Quality control

Germline DNA was genotyped at the Affymetrix Research Services Laboratory (Santa Clara, CA) using default quality control (QC) and genotype calling criteria. Variants failed QC if: (1) the call rate was < 95%; (2) p-values of Hardy-Weinberg equilibrium in controls were < 10^{-7}; or (3) there was > 2% discordance in duplicate pairs. Further, monomorphic variants were removed. Of 6,948 variants genotyped within 11 known EOC risk regions, 4,919 (71%) met these QC criteria and were polymorphic. Outside of these regions, of 7,189 variants selected based on whole-genome sequencing data, 5,286 (74%) met QC criteria and were polymorphic. Of 1,302 variants associated with NF-κB binding, 980 (75%) met QC criteria and were polymorphic, and, of 2,000 variants selected from a GWAS of endometrioid EOC (N = 2,000), 1,826 (91%) met QC criteria and were polymorphic. Most variants were excluded for being monomorphic. Thus, a total of 13,011 genotyped variants remained for analysis.

Association analysis

All cases were included in the overall EOC risk association analyses (N=4,973). Subset analyses were performed on histologic subsets based on a priori selection; HGSC (N=3,573) and endometrioid EOC (N=835). For each analysis, 5,640 controls were used. Associations were estimated using logistic regression assuming an additive genetic model, adjusting for age, study site, and population substructure by including the first three eigenvalues from principal components analysis[18]. All analyses were conducted in R version 3.0.2 (http://www.R-project.org/).

ACKNOWLEDGMENTS

The NHS/NHSII studies thank the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

GRANT SUPPORT

Funding support for the Follow-up of Ovarian Cancer Genetic Association and Interaction Studies (FOCI) was provided through the National Cancer Institute’s Cancer Post-GWAS Initiative, Genetic Associations and Mechanisms in Oncology (GAME-ON) (U19-CA148112). In addition, we acknowledge the following: DOV: National Institutes of Health R01-CA112523 and R01-CA87538; HAW: National Institutes of Health (R01-CA58598, N01-CN-55424 and N01-PC-67001); HOP: DOD DAMD17-02-1-0669 and NCI K07-CA080668, R01-CA95023, P50-CA159981; NIH/National Center for Research Resources/General Clinical Research Center grant M01-RR000056; MAY: National Institutes of Health (R01-CA122443, P30-CA15083-41, P50-CA136393); NCO: National Institutes of Health (R01-CA76016) and the Department of Defense (DAMD17-02-1-0666); NEC: National Institutes of Health (R01-CA154419 and P50-CA105009) and Department of Defense W81XWH-10-1-02802; NIH: NIH (P01-CA87696 and R01-CA94494); NJO: National Cancer Institute (K07-CA095666, R01-CA83918, K22-CA138563, and P30-CA072720) and the Cancer Institute of New Jersey; NCI CCSG award (P30-CA008748); OVA: This work was supported by Canadian Institutes of Health Research grant (MOP-86727); LEK is supported by a Canadian Institutes of Health Research New Investigator award (MSH-87734); POL: Intramural Research Program of the National Cancer Institute; UCI: R01-CA058860, R01-CA092044, US Public Health Service PSA-042205, and the Lon V Smith Foundation grant LVS-39420; USC: P01-CA17054, P30-CA14089, R01-CA61132, N01-PC67010, R03-CA113148, R03-CA115195, N01-CN025403, and California Cancer Research Program (00-01389V-20170, 2II0200). CMP is supported by National Institutes of Health R01-CA-149429.

REFERENCES


