Investigation of exomic variants associated with overall survival in ovarian cancer

Stacey J Winham*[1], Ailith Pirie*[2], Yian Ann Chen[3], Melissa C Larson[1], Zachary C Fogarty[1], Madalene A Earp[4], Hoda Anton-Culver[5], Elisa V Bandera[6], Daniel Cramer[7,8], Jennifer A Doherty[9], Marc T Goodman[10], Jacek Gronwald[11], Beth Y Karlan[12], Susanne K Kjaer[13,14], Douglas A Levine[15], Usha Menon[16], Roberta B Ness[17], Celeste L Pearce[18], Tanja Pejovic[19,20], Mary Anne Rossing[21,22], Nicolas Wentzensen[23], Yukie T Bean[19,20], Maria Bisogna[15], Louise A Brinton[23], Michael E Carney[24], Julie M Cunningham[25], Cezary Cybulski[26], Anna deFazio[27,28], Ed M Dicks[29], Robert P Edwards[30], Simon A Gayther[18], Aleksandra Gentry-Maharaj[16], Martin Gore[31], Edwin S Iversen[32], Allan Jensen[13], Sharon E Johnatty[33], Jenny Lester[12], Hui-Yi Lin[3], Jolanta Lissowska[34], Jan Lubinski[11], Janusz Menkiszak[35], Francesmary Modugno[30,36,37], Kirsten B Moysich[38], Irene Orlow[39], Malcolm C Pike[18,39], Susan J Ramus[18], Honglin Song[29], Kathryn L Terry[7,8], Pamela J Thompson[10], Jonathan P Tyrer[29], David J van den Berg[18], Robert A Vierkant[1], Allison F Vitonis[7], Christine Walsh[12], Lynne R Wilkens[40], Anna H Wu[18], Hannah Yang[23], Argyrios Ziogas[41], Andrew Berchuck[42]; Georgia Chenevix-Trench on behalf of Australian Ovarian Cancer Study Group[28,33,43,44]; Joellen M Schildkraut[45], Jennifer Permuth-Wey[3], Catherine M Phelan[3], Paul D P Pharoah[29], Brooke L Fridley[47], Thomas A Sellers#[3], and Ellen L Goode#[4]

1. Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA.
2. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
3. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.
4. Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, 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. Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
8. Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.
9. Section of Biostatistics and Epidemiology, The Geisel School of Medicine at Dartmouth, Lebanon, NH, USA.
10. Samuel Oschin Comprehensive Cancer Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA.
11. Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland.
12. Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
13. Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark.
14. Department of Gynecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.
15. Gynecology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY, USA.
16. Gynaecological Cancer Research Centre, Department of Women's Cancer, Institute for Women's Health, UCL, London, UK.
17. The University of Texas School of Public Health, Houston, TX, USA.

18.   Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
19.   Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA.
20.   Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA.
21.   Department of Epidemiology, University of Washington, Seattle, WA, USA.
22.   Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
23.   Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.
24.   John A Burns School of Medicine, Obstectrics and Gynecology, University of Hawaii, Honolulu, HI, USA.
25.   Department of Laboratory Medicine and Pathology, Division of Experimental Pathology, Mayo Clinic, Rochester, MN, USA.
26.   International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical Academy, Szczecin, Poland.
27.   Department of Gynaecological Oncology, Westmead Hospital, Sydney, Australia.
28.   Center for Cancer Research, University of Sydney at Westmead Millennium Institute, Sydney, Australia.
29.   Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
30.   Department of Obstetrics, Gynecology and Reproductive Sciences, Division of Gynecologic Oncology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.
31.   Gynecological Oncology Unit, The Royal Marsden Hospital, London, UK.
32.   Department of Statistical Science, Duke University, Durham, NC, USA.
33.   Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
34.   Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center & Institute of Oncology, Warsaw, Poland.
35.   Department of Surgical Gynecology and Gynecological Oncology of Adults and Adolescents, Pomeranian Medical University, Szczecin, Poland.
36.   Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA.
37.   Womens Cancer Research Program, Magee-Women's Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA.
38.   Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA.
39.   Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA.
40.   Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA.
41.   Department of Epidemiology, Center for Cancer Genetics Research and Prevention, School of Medicine, University of California Irvine, Irvine, CA, USA.
42.   Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA.
43.   Cancer Division, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia.
44.   Peter MacCallum Cancer Institute, Melbourne, VIC, Australia.
45.   Public Health Services, University of Virginia, Charlottesville, Virginia, USA.
46.   Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core, University of Kansas Cancer Center, Kansas City, KS, USA.

*Equal contributions

[#]Equal contributions

Running Title: Exomic variants associated with ovarian cancer survival

Keywords: epithelial ovarian cancer, overall survival, exome

Correspondence: Ellen L. Goode, Ph.D., M.P.H., Department of Health Sciences, Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA, Phone 507/266-7997, Fax 507/266-2478, Email egoode@mayo.edu

Conflict of Interest: The authors have no conflicts of interest to disclose.

Word Count: 4,228

Number of Tables: 3

Number of Figures: 3

Abstract

Background: While numerous susceptibility loci for epithelial ovarian cancer (EOC) have been identified, few associations have been reported with overall survival. In the absence of common prognostic genetic markers, we hypothesize that rare coding variants may be associated with overall EOC survival and assessed their contribution in two exome-based genotyping projects of the Ovarian Cancer Association Consortium (OCAC).

Methods: The primary patient set (Set 1) included 14 independent EOC studies (4293 patients) and 227,892 variants, and a secondary patient set (Set 2) included six additional EOC studies (1744 patients) and 114,620 variants. Because power to detect rare variants individually is reduced, gene-level tests were conducted. Sets were analyzed separately at individual variants and by gene, and then combined with meta-analyses (73,203 variants and 13,163 genes overlapped).

Results: No individual variant reached genome-wide statistical significance. A SNP previously implicated to be associated with EOC risk and, to a lesser extent, survival, rs8170, showed the strongest evidence of association with survival and similar effect size estimates across sets (P_{meta}=1.1E-6, HR_{Set1}=1.17, HR_{Set2}=1.14). Rare variants in *ATG2B*, an autophagy gene important for apoptosis, were significantly associated with survival after multiple testing correction (P_{meta}=1.1E-6; $P_{corrected}$=0.01).

Conclusions: Common variant rs8170 and rare variants in *ATG2B* may be associated with EOC overall survival, although further study is needed.

Impact: This study represents the first exome-wide association study of EOC survival to include rare variant analyses, and suggests that complementary single variant and gene-level analyses in large studies are needed to identify rare variants that warrant follow-up study.

Introduction

Epithelial ovarian cancer (EOC) is diagnosed in over 230,000 women world-wide every year and it is a leading cause of cancer death (1). Most women are diagnosed with advanced stage disease, when five-year survival rates are low (2, 3). Several clinical and demographic factors are associated with survival, such as stage, grade and histological subtype (4), but few germline prognostic variants have been identified (5). The strongest known genetic risk factors for EOC are rare *BRCA1* and *BRCA2* mutations (6), which also confer differences in clinical characteristics including improved five-year, but not ten-year, survival (7-9). Although genome-wide association studies (GWAS) to identify common risk loci have been fruitful (10-15), studies of survival have been less so (5). For example, variant rs8170 at the 19p13 locus—a region associated with ovarian and breast cancer risk phenotypes and known to interact with *BRCA1* (10, 16)— was associated with EOC survival in an initial phase of a GWAS (HR=1.35, P=2.4E-4) conducted by the Ovarian Cancer Association Consortium (OCAC), but not replicated in an independent dataset (HR=1.01, P=0.85) (10).

While the lack of identified common germline prognostic markers from earlier GWAS may be due to inadequate power (due to limited sample size or study heterogeneity), it is also possible that rare variants (in addition to those in *BRCA1* and *BRCA*2) not captured in GWAS arrays may be associated with overall EOC survival. In fact, it has been suggested that multiple rare variants of large effect could collectively be responsible for some of the 'missing heritability' not explained by the common variants of modest effect identified through GWAS (17). Motivated by this rare variant hypothesis (18), commercial genotype arrays based on exome sequencing studies that attempted to capture all variants (common and rare) within coding regions have emerged as a new approach. As single marker tests have very little power to detect association with rare variants, approaches that pool information across all variants within a gene region may provide improved power (19-23). We hypothesize that rare variants, either

individually or collectively across a gene, may be associated with overall survival in EOC. To test this hypothesis, we combined data from two exome-based OCAC genotyping projects that used commercial arrays based on 16 major exome sequencing projects (24). With over 6,000 EOC patients, this is the first exome-wide rare variant assessment of genetic associations with EOC overall survival.

Materials and Methods

Study Participants

Study participants included 20 independent studies of EOC (Supplemental Table S1), which consisted of two sample sets that were genotyped separately on different platforms.

Set 1: For Set 1, 14 independent studies of EOC (DOV, HAW, HOP, LAX, MAC, MAY, MSK, NCO, NEC, NJO, ORE, POL, UCI, USC) served as the primary sample (Supplemental Table S1), including 6293 patients. Patients consisted of women aged 18 and older with a pathologically confirmed primary invasive EOC, fallopian tube cancer, or primary peritoneal cancer, excluding Brenner tumors or those missing tumor histology. Patients were excluded with incomplete survival time information, such as missing vital status or time from diagnosis to follow-up, resulting in 4,976 patients. To avoid potential issues with population stratification, only patients of European ancestry were analyzed (N=4293; 2257 deaths at 10 years post-diagnosis).

Set 2: The smaller Set 2 consisted of six additional independent studies of invasive EOC (AUS, MAL, POC, RMH, SEA, UKO) (Supplemental Table S1) with 1878 patients of European ancestry. To increase statistical power, Set 2 was combined with Set 1 via meta-analysis (due to separate genotyping). Patients with incomplete survival time information (e.g. missing vital status, enrollment greater than 10 years after diagnosis) were excluded, resulting in 1744 patients available for analysis (1027 deaths at 10 years post-diagnosis). Sixteen patients had missing follow-up time, and were set to

7

the median follow-up time of the study; six patients had incomplete time-to-entry information, and were set to entry at time zero. Characteristics of both sample sets are described in Table 1. Both patient sets consist of predominantly patients with high-grade serous histology, although Set 1 also includes a large number of endometrioid patients.

Genotyping and Quality Control

For Set 1, genotyping was performed using genomic DNA and whole genome amplified (WGA) DNA for 5904 and 389 patients, respectively. Genotyping was performed at Affymetrix Corporation using the Affymetrix Axiom Exome Array (www.affymetrix.com) of 409,582 markers, including standard content markers, as well as 100,000 custom markers chosen based on preliminary survival associations, follow-up of prior risk associations, and non-coding candidate gene variants. The default quality control (QC) criteria recommended by Affymetrix was applied to all samples, and then the WGA and genomic samples were processed separately. For genomic samples, genotypes were recalled using Powertool for variants with MAF < 5%, and samples and variants with call rates < 97% were excluded. WGA samples and variants with call rates < 97% were also excluded and then merged with genomic samples, resulting in 384,029 variants. Excluding monomorphic variants, 227,892 variants were used in analysis.

For Set 2, patients contributed genomic DNA samples that were genotyped at the University of Cambridge on the Illumina Infinium HumanExome BeadChip (www.Illumina.com) at a total of 247,840 exonic markers. Genotype calling was carried out according to Best Practice Guidelines using the GenCall module in Illumina's Genome Studio with a default GenCall threshold of 0.15 (25). Samples with low call rate (<99%), high or low heterozygosity based on common SNPs (MAF ≥ 0.05), and ambiguous sex or relatedness were excluded. A total of 130,909 variants were monomorphic and 147 variants were excluded due to low call rate, resulting in 114,620 markers.

Of the 73,203 variants overlapping on both genotyping arrays, a majority are rare variants (median MAF=0.0028, IQR=0.068) and coding variants (89.2%). Supplemental Figure S1 displays the MAF distribution by functional category.

Single Variant Analysis

Both patient sets were analyzed using Cox proportional hazards regression to estimate the association of each variant with overall 10-year survival time. For each variant, a Cox proportional hazards model was fit to the number of copies of the minor allele, and a likelihood ratio test was conducted, accounting for left truncation (26) and right-censoring after 10 years. For top variants, Kaplan-Meier curves were also generated that account for left truncation and right censoring. Because the primary and secondary patient sets were genotyped separately and on different platforms, they were analyzed separately.

For the primary patient set, models were first minimally adjusted for age, site, and three principal components (PCs) to account for population stratification. PCs were computed based on common variants (MAF > 1%, HWE p > 1E-7) using the study patients, as well as HapMap subjects. Additionally, the following clinical factors with univariate survival associations of p<0.05 and with p<0.0001 in multivariable modelling were included as covariates along with three PCs: age, stage, grade, histology, and site. Both sample sets consisted of patients of primarily serous histology, although Set 1 also included many endometrioid cases. Therefore all histologies were included in analyses (N=4293), but histological subtype-specific analyses were also conducted for high-grade serous (N=3149) and endometrioid subtypes (N=735), because some candidate variants were targeted for these subtypes. Variants with p < 2.2E-7 were considered significant after Bonferroni correction for multiple testing.

Because the sample size of Set 2 was relatively small, its primary utility was to improve power and provide supportive evidence for the primary results via meta-analysis; the patient sets were

9

genotyped separately, which limited the ability to conduct a pooled analysis. For Set 2, Cox proportional hazards analyses (accounting for left truncation and right-censoring after 10 years) were conducted using minimal adjustments as above; no other covariates were associated with the outcome and therefore additional adjustments were not included due to the smaller sample size. Fixed effect meta-analyses were conducted for the overlapping variants on both the Affymetrix and Illumina platforms (73,203 variants), based on the likelihood ratio summary statistics from the minimally adjusted Set 1 and Set 2 analyses. Single variant meta-analysis was conducted using the R package 'rMeta' (http://cran.r-project.org/web/packages/rmeta/). All analyses were conducted in R version 3.0.2 (http://www.R-project.org/).

Gene-Level Analysis

As single variant analyses are under-powered for rare variants, gene-level tests were also used. Variants were mapped to genes based on Human Genome Build 37 and annotation supplied by Affymetrix and Illumina, resulting in 18,323 genes analyzed in Set 1 and 13,191 genes analyzed in Set 2. To be included in gene-level analysis, genes were required to contain at least two SNPs. Gene-level testing was conducted on both the sample sets separately using burden tests, as well as Sequence Kernel Association Test (SKAT) for Cox proportional hazards models (27). The gene-level burden test is a weighted sum of the genotypes of all variants with a gene regressed on survival time (20). The gene-level SKAT statistic is a weighted sum of the single variant likelihood ratio test statistics, across all variants within a gene (23). Variant weights were computed as a function of MAF, where all variants were included, and rare variants received much greater weight; specifically, the weight function used was $\sqrt{w_j}$=Beta(MAF$_j$,1,25) for each variant j (23). Minimally adjusted (Sets 1 and 2) and fully-adjusted tests (Set 1) were computed as described above for single variant tests. Gene-level tests (SKAT and burden tests) were computed using the R package 'seqMeta' (http://cran.r-project.org/web/packages/seqMeta/), with minor modifications to accommodate left-truncated data.

As with the single variant tests, meta-analyses of gene-level tests across the sample sets were also conducted. For the purpose of meta-analysis only, variants that were monomorphic in one set but not the other were included to provide information on MAF across both studies. A total of 13,163 genes available in both sample sets were meta-analyzed using the R-package 'seqMeta'. Genes with p<3.8E-6 were considered significant after Bonferroni correction for multiple testing. Sensitivity analyses were conducted for variants with MAF<0.05, MAF<0.10, and MAF<0.25, as well as for coding variants only.

Results

Single variants

Results for all 73,203 variants included in the meta-analysis of 6,054 invasive EOC patients are displayed in Figure 1. Three variants were associated with overall survival at p-values < 1.0E-4, including two with MAF < 0.001 (Table 2). The top ranking variant in the meta-analysis was common variant rs8170 in *BABAM1* at 19p13.11 (P_{meta}=1.14E-6), previously reported to be associated with EOC risk (10); the minor allele of this variant was associated with decreased survival in both sets (HR_1=1.17, HR_2=1.14; Figure 2). Additional top ranking variants were rare non-synonymous variants rs140079492 in *PAK7* (P21 protein (cdc42/rac)-activated kinase) at 20p12.2 ((P_{meta}=9.32E-6) and rs34335714 in *CTSW* (cathepsin W) at 11q13.1 (P_{meta}=4.91E-5, Table 2).

Despite reduced sample size compared to meta-analysis, Set 1-specific results are also of interest due to the much larger number of variants targeted than in Set 2 (including custom candidate variants); therefore we also report the results of all variants analyzed in Set 1 (including those non-overlapping with Set 2). In this primary EOC sample set (4293 patients), with adjustment for age, site, stage, grade, histology, and PCs, 24 variants in 11 chromosomal regions showed evidence of association with overall survival with P<5.0E-5, although none of these associations are significant after correction

for multiple testing. Genome-wide results are plotted in Supplemental Figure S2, and top variants are described in Supplemental Table S2. The variant most strongly associated with survival was rs7642051, a common variant (MAF=0.46) in *LMCD1-AS1* (LMCD1 antisense RNA 1) at 3p26.1 (p=4.10E-6); its minor allele was associated with decreased survival (HR=1.15; Supplemental Table S2, Supplemental Figure S3). Several correlated variants in an intergenic region of chromosome 1 between *UBE2U* and *CACHD1* (1p31.3), targeted due to a prior study of overall survival, also showed modest evidence for association with survival (Supplemental Figure S4). This region showed similar results for the high grade serous subtype (data not shown). Among patients with endometrioid subtype, the most strongly associated variant was rs757759 in *TBXAS1* at 7q34 (HR=1.78, P=1.55E-6, MAF=0.23).

Gene-level

Results for all 13,163 genes included in the meta-analysis across Sets 1 and 2 are displayed in Figure 3. Five genes had meta-analysis p<1.0E-4 based on the burden test, and four of these genes showed evidence of association in both sets separately (Table 3). The top-ranking gene was *ATG2B* (P_{meta}=1.08E-6), which was significant after multiple testing correction based on 13,163 genes ($P_{corrected}$=0.014). Out of 17 single variants meta-analyzed in *ATG2B* at 14q32.2, nine had p<0.10 (Supplemental Figure S5) and 15 with positive effect size estimates (HR>1.0); of note, variants were largely uncorrelated (pairwise r^2 <0.2), and most were rare (16 with MAF<0.01; median=5.9E-4, range=1.7E-4 – 6.8E-3). When restricted to the high-grade serous subset, *ATG2B* remains the top gene (P_{meta}=1.75E-6). Other genes of interest across all histologies include zinc finger protein *PEG3* (P_{meta}=1.82E-5), helicase-like transcription factor *HLTF* (P_{meta}=1.9E-5), mitochondrial fission regulator 1-like gene *MTFR1L* (p=3.1E-5), and T-cell surface glycoprotein *CD1E* (P_{meta}=8.78E-5; Table 3). Results differed based on the SKAT meta-analysis; while gene-rankings were relatively similar to the burden test, p-values were higher (min P=1.17E-4 for *ZNF131*). In particular, the meta-analysis p-value based on SKAT

for *ATGB2* was p=0.006 (Table 3). Although top-ranked in single variant analyses, *BABAM1* was not significant at the gene-level (P_{burden}=0.83, P_{SKAT}=0.53).

In analysis of Set 1 alone, when adjusted for age, site, stage, grade, histology, and PCs, five genes showed evidence of association with overall survival at $P < 1.0 \times 10^{-4}$ from the burden test, although none of these associations are significant after multiple testing correction based on number of genes (Supplemental Table S3); results were similar using SKAT (Supplemental Figure S6). The top-ranked genes were *POGLUT1* (protein O-glucosyltransferase 1; P_{burden}=7.10E-6, P_{SKAT}=8.33E-6), *ST20* (suppressor of tumorigenicity 20; P_{burden}=2.32E-5, P_{SKAT}=2.32E-5), and *ATG2B* (autophagy related 2B; P_{burden}=5.16E-5, P_{SKAT}=3.23E-2)—the most significant gene in meta-analysis. Additionally, consistent with individual variant analysis, variation in *UBE2U* showed some evidence of association (36 variants, P_{burden}=5.07E-4, P_{SKAT}=4.95E-4), although evidence was less strong in the *CACHD1* region (65 variants, P_{burden}=0.004, P_{SKAT}=0.025). Gene-level results were similar when restricted to patients with high-grade serous subtype, when restricted to variants with MAF<0.05, MAF<0.10, and MAF<0.25, and when restricted to only coding variants (data not shown); ATG2B remained the top gene, with meta-analysis p-values based on a burden test ranging from 3.94E-7 to 1.08E-6.

Discussion

To our knowledge, this study represents the first exome-wide assessment of association between EOC survival and rare genetic variation, both for single variants and for genes, including over 6000 patients. By combining data from two exome-based OCAC genotyping projects, genome-wide significant findings were identified at the gene-level in *ATG2B* (P_{meta}=1.1E-6, $P_{corrected}$=0.01); and, in analysis of approximately 73,000 variants, the known EOC susceptibility variant rs8170 in *BABAM1* arose as the most statistically significant (P_{meta}=1.1E-6), followed by non-synonymous variants rs140079492 in

PAK7 (serine/threonine-protein kinase; missense mutation Glu → Gly) and rs34335714 in *CTSW*

(cathepsin W; missense mutation Ser → Asn, predicted splice variant). As genotype imputation methods

are known to be unreliable in the context of rare variation (28, 29), the use of direct genotyping is an

important strength of this report.

  ATG2B, the survival-associated gene based on burden testing, is a key autophagy gene residing

on chromosome 14q32.2. Autophagy is a cell death process which uses degradation of lysosomal

cytoplasmic components; in cancer, it is thought to link apoptosis with programmed necrosis and has

been proposed as an alternative target to treat tumor resistance (30). It is regulated by several

microRNAs, including mir-30d and mir-130a (31, 32), and has recently been linked to precursors of

pancreatic cancer (33).   Here, up to 34 *ATG2B* variants genotyped in either Sets 1 or 2 associated in

combination with EOC survival; this was consistent with the trends of individual *ATG2B* variants, where

nine of 17 *ATG2B* variants observed in both sets had p<0.10 in the meta-analysis. These results suggest

that *ATG2B* may warrant targeted sequencing in large datasets to confirm the existence of potentially

rare, inherited prognostic variants.

  Although not statistically significant after correction for multiple testing, *BABAM1* rs8170 at

19p13, the top ranking individual variant across both sample sets (adjusting for age, site, and PCs),

warrants attention. This synonymous coding variant is relatively common (MAF=0.20) and is included

on multiple commercial GWAS arrays, as well as the Affymetrix and Illumina exome-based arrays used

here. Its MAF contributed to increased power to its detection over the rarer variants on the arrays.

BABAM1 rs8170 is known to associate with EOC risk (10), with triple-negative breast cancer risk (34),

and with risk of ovarian cancer and estrogen-negative subtypes of breast cancer among *BRCA1* and

BRCA2 mutation carriers (16, 35). The estimated hazard ratio in meta-analysis was 1.16, which roughly

translates to a 16% reduction in median survival time for patients with one copy of the minor allele

compared to those with no copies (36).   Sensitivity analysis of the current data show that when also

14

adjusted for stage, grade, and histology, the rs8170 effect size was slightly reduced (HR=1.11 vs. HR=1.17; P=4.33E-5). However, when the analysis was restricted to patients with high-grade serous cancer, effect size estimates were relatively robust whether adjusted for age, site, and PCs (HR=1.17, P=4.66E-5) or also stage and grade (HR=1.14, P=8.02E-5). Additional patient subset analyses may uncover other patterns of survival; for example, for a subset of OCAC with available data regarding detailed treatment strategies, no association of rs8170 with overall survival or progression free survival was observed after stratifying by treatment (37). Although prior EOC survival studies yielded inconsistent results (10), unpublished meta-analysis of genotyped and imputed variation at rs8170 from a much larger sample of over 18,000 EOC patients (including the majority of those in the current study and all of the patients from a prior EOC survival study(10)) revealed a genome-wide significant association (HR=1.15, P=4.7E-9) (Pharoah et al, personal communication). Altogether, the *BABAM1* region containing rs8170 is clearly of importance to ovarian cancer and will be subjected to fine-mapping and detailed functional analyses.

As one of the early GWAS using exome-based Affymetrix and Illumina genotyping arrays, we note some rare variant analytical lessons learned. Single variant analyses often yielded different results than the aggregate gene-level tests, and represent useful, complementary analysis strategies. Because single variant tests were subject to reduced power to detect associations with rare variants, only associations with common variants (such as rs8170) and rare variants with extremely large effect sizes were detectable. For example, in the single variant meta-analysis, the median MAF is less than 0.005, which translates to less than 60 subjects out of the combined Set 1 and Set 2 sample size of 6037 that can be expected to harbor at least one copy of the rare allele. For rare variants with MAF of 0.005, we have 80% power to detect hazard ratios greater than 2.9, and for even rarer variants, only much larger hazard ratios would be detectable with this sample size. Thus, larger sample sets will be necessary to rule out or detect individual associations with the approximately 70,000 variants that were individually

meta-analyzed here. However, because gene-level tests consider all variants within a region and weight rare variants more heavily, they are able to identify associations with regions that contain many variants with moderate evidence of association; for example, no variant in *ATG2B* had a meta-analysis p-value<0.001, yet it was genome-wide significant at the gene-level.

The most powerful gene-level analysis tool is dependent on the underlying genetic etiology of the trait, and because the true genetic architecture of EOC survival is unknown, we used two methods for gene-level testing that have been shown to be powerful under different scenarios(20, 27). While the gene rankings using a burden test or SKAT were similar, the p-values based on the SKAT analysis were often higher, suggesting that SKAT may be less powerful, consistent with simulation studies (23). The burden test assumes that the direction of effect for each variant within the gene region is the same, while SKAT does not assume consistent effect direction. Therefore the burden test will be more powerful to detect rare variants if the true causal effects are in the same direction (23); because our outcome of interest in this study was survival, it is not unreasonable to assume that rare variants would increase risk of death (rather than the converse), although of course the true underlying genetic model of EOC survival remains unknown. In fact, for *ATG2B* (where the SKAT p-value was higher than the burden p-value), all but one of the nine variants with meta-p<0.10 had estimated hazard ratios greater than 1.0.

The findings of this study should be considered in the context of the following limitations, in addition to sample size. The two patient sets were genotyped separately on different platforms that targeted different variants due to differences in chemistry (Affymetrix vs. Illumina), even though both panels were designed from a common set of variants from the Exome Sequencing Project (24). This is far from ideal, not only requiring extensive additional data harmonization measures, but limiting our ability to conduct a pooled analysis and restricting meta-analysis to a reduced set of variants (only

73,203 combined from 227,892 Set 1 and 114,620 Set 2 variants) and genes (only 13,163 combined from

18,323 Set 1 and 13,191 from Set 2 genes) that were observed in both sets. Therefore many rare

variants that were observed in the larger Set 1 were not assessed in Set 2. Furthermore, although also a

strength, approximately 25% of the Affymetrix array was customized content, and therefore not

interrogated in Set 2, including the intergenic region in chromosome 1 between *UBE2U* and *CACHD1*.

Imputation to a denser set of variants, for example using 1000 Genomes Project data, is known to be

unreliable based on rare variants and was therefore not performed; future work to integrate these data

with other available genotypes for these patients will be informative. Additionally, while all patients

underwent surgery, the important clinical factors of optimal debulking and type of treatment (e.g.

chemotherapy) could not be included as covariates or stratification factors due to a large amount of

missing data; both should be examined in future genetic studies to assess their impact on survival

outcomes. Finally, it should be noted that some of the data analyzed here were not independent from

previous OCAC EOC overall survival reports(5, 10, 38); specifically, some variants on the exome arrays

were also on GWAS or candidate gene arrays which had previously been used on a subset of the current

patients. However, the vast majority of the current array content was largely used to test the novel rare

variant hypothesis.

Lastly, this study comprised data from 20 OCAC studies, with differing ascertainment strategies

and methods for clinical follow-up of cases, resulting in the potential for increased study heterogeneity

in the overall survival outcome measure. While we adjusted for study site in all models, uncaptured

heterogeneity across studies (such as differences in treatment strategies over time and across medical

institutions) will lead to increased noise and reduced power. In future studies, more refined outcome

measures targeted at reducing this heterogeneity could improve power. For example, because first-line

therapy may differ from treatment strategies after a recurrence, progression free survival may be a

more optimal endpoint, although this measure is also susceptible to study site heterogeneity. Because

recurrence data was not widely available for all patients across all sites, it was not considered as a primary endpoint in this study. However, analysis of the association between rs8170 and progression free survival were consistent with the results observed for the overall survival outcome (HR_{meta}=1.15, P_{meta}=1.3E-6, 95% CI=1.09-1.22).

To summarize, in the absence of associations between common genetic variants and EOC survival, we examined the role of primarily rare exonic variants, individually and collectively across genes, using partially customized commercial arrays in over 6000 patients. While no individual variants were significant at the genome-wide level, *ATG2B* showed gene-level evidence of association with overall survival, and *BABAM1* rs8170 was again highlighted to be of particular relevance. This suggests potential candidates for future studies that may lead to targets for improved EOC outcomes, although follow-up in the larger sample of OCAC subjects not included here, as well as a focus on reducing the heterogeneity in the outcome measure, will be critical.

Acknowledgements

References

1.	Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136:E359-86.

2.	McGuire V, Jesser CA, Whittemore AS. Survival among U.S. women with invasive epithelial ovarian cancer. Gynecol Oncol. 2002;84:399-403.

3.	Howlader N, Noone A, Krapcho M, Garshell J, Miller D, Altekruse S, et al. SEER cancer statistics review, 1975-2011. Bethesda, MD: National Cancer Institute; 2014.

4.	Kosary CL. FIGO stage, histology, histologic grade, age and race as prognostic factors in determining survival for cancers of the female gynecological system - an analysis of 1973-87 seer cases of cancers of the endometrium, cervix, ovary, vulva, and vagina Seminars in Surgical Oncology. 1994;10:31-46.

5.	White KL, Vierkant RA, Fogarty ZC, Charbonneau B, Block MS, Pharoah PDP, et al. Analysis of over 10,000 cases finds no association betweenpreviously reported candidate polymorphisms and ovarian cancer outcome. Cancer Epidemiology Biomarkers & Prevention. 2013;22:987-92.

6.	Easton DF, Bishop DT, Ford D, Crockford GP. Genetic-linkage analysis in familial breast and ovarian - cancer- results from 214 - families. American Journal of Human Genetics. 1993;52:678-701.

7.	Cass I, Baldwin RL, Varkey T, Moslehi R, Narod SA, Karlan BY. Improved survival in women with BRCA-associated ovarian carcinoma. Cancer. 2003;97:2187-95.

8.	Bolton KL, Chenevix-Trench G, Goh C, Sadetzki S, Ramus SJ, Karlan BY, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. Journal of the American Medical Association. 2012;307:382-90.

9.	Candido-Dos-Reis FJ, Song H, Goode EL, Cunningham JM, Fridley BL, Larson MC, et al. Germline Mutation in BRCA1 or BRCA2 and Ten-Year Survival for Women Diagnosed with Epithelial Ovarian Cancer. Clin Cancer Res. 2015;21:652-7.

10.	Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nature Genetics. 2010;42:880-+.

11.	Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nature Genetics. 2010;42:874-+.

12.	Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. 2013;45:362-70, 70e1-2.

13.	Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet. 2009;41:996-1000.

14.	Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nat Genet. 2015.

15.	Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature Genetics. 2013;45:371-84, 84e1-2.

16.	Antoniou AC, Wang XS, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nature Genetics. 2010;42:885-+.

17.	Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics. 2010;11:446-50.

18.   Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics. 2008;40:695-701.
19.   Li BS, Leal SM. Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. American Journal of Human Genetics. 2008;83:311-21.
20.   Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. Plos Genetics. 2009;5.
21.   Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genetic Epidemiology. 2010;34:188-93.
22.   Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, et al. Testing for an unusual distribution of rare variants. Plos Genetics. 2011;7.
23.   Wu MC, Lee S, Cai TX, Li Y, Boehnke M, Lin XH. Rare-variant association testing for sequencing data with the sequence kernel association test. American Journal of Human Genetics. 2011;89:82-93.
24.   Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013;493:216-20.
25.   Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques. 2002;Suppl:56-8, 60-1.
26.   Andersen PK, Gill RD. Cox regression- model for counting- processes- a large samply study Annals of Statistics. 1982;10:1100-20.
27.   Chen H, Lumley T, Brody J, Heard-Costa NL, Fox CS, Cupples LA, et al. Sequence kernel association test for survival traits. Genetic Epidemiology. 2014;38:191-7.
28.   Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95:5-23.
29.   Li L, Li Y, Browning SR, Browning BL, Slater AJ, Kong X, et al. Performance of genotype imputation for rare variants identified in exons and flanking regions of genes. PLoS One. 2011;6:e24945.
30.   Nickells RW. The cell and molecular biology of glaucoma: mechanisms of retinal ganglion cell death. Invest Ophthalmol Vis Sci. 2012;53:2476-81.
31.   Yang X, Zhong X, Tanyi JL, Shen J, Xu C, Gao P, et al. mir-30d Regulates multiple genes in the autophagy pathway and impairs autophagy process in human cancer cells. Biochem Biophys Res Commun. 2013;431:617-22.
32.   Kovaleva V, Mora R, Park YJ, Plass C, Chiramel AI, Bartenschlager R, et al. miRNA-130a targets ATG2B and DICER1 to inhibit autophagy and trigger killing of chronic lymphocytic leukemia cells. Cancer Research. 2012;72:1763-72.
33.   Permuth-Wey J, Chen YA, Fisher K, McCarthy S, Qu X, Lloyd MC, et al. A Genome-Wide Investigation of MicroRNA Expression Identifies Biologically-Meaningful MicroRNAs That Distinguish between High-Risk and Low-Risk Intraductal Papillary Mucinous Neoplasms of the Pancreas. PLoS One. 2015;10:e0116869.
34.   Stevens KN, Fredericksen Z, Vachon CM, Wang X, Margolin S, Lindblom A, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. Cancer Research. 2012;72:1795-803.
35.   Couch FJ, Gaudet MM, Antoniou AC, Ramus SJ, Kuchenbaecker KB, Soucy P, et al. Common variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. Cancer Epidemiology, Biomarkers & Prevention. 2012;21:645-57.
36.   Cortes J, Gonzalez JA, Campbell MJ, Cobo E. A hazard ratio was estimated by a ratio of median survival times, but with considerable uncertainty. J Clin Epidemiol. 2014;67:1172-7.
37.   Johnatty S, Tyrer JP, Kar SP, Beesley J, Lu Y, Gao B, et al. Genome-wide analysis identifies novel loci associated with ovarian cancer outcomes: findings from the Ovarian Cancer Association Consortium. Clin Cancer Res. 2015.

38.   Fridley BL, Jenkins GD, Tsai YY, Song H, Bolton KL, Fenstermacher D, et al. Gene set analysis of survival following ovarian cancer implicates macrolide binding and intracellular signaling genes. Cancer Epidemiol Biomarkers Prev. 2012;21:529-36.

Tables

Table 1: Characteristics of EOC sample sets.

	Sample Set 1			Sample Set 2		
	Alive (N=2036)	Deceased (N=2257)	Total (N=4293)	Alive (N=717)	Deceased (N=1027)	Total (N=1744)
Median Survival Time (years, 95% CI)	4.91 (4.65,5.21)	2.56 (2.44,2.68)	4.58 (4.41,4.77)	6.03 (5.70,6.43)	2.89 (2.69,3.08)	4.32 (4.10,4.60)
Person Years (years)	10,737.46	6,169.05	16,906.51	3,545.56	2,805.42	6,350.98
Grade						
Well differentiated	288 (14.1%)	79 (3.5%)	367 (8.5%)	69 (9.6%)	72 (7.0%)	141 (8.1%)
Moderately differentiated	426 (20.9%)	370 (16.4%)	796 (18.5%)	109 (15.2%)	216 (21.0%)	325 (18.6%)
Poorly differentiated	886 (43.5%)	1318 (58.4%)	2204 (51.3%)	3237 (47.0%)	554 (53.9%)	891 (51.1%)
Undifferentiated*	276 (13.6%)	302 (13.4%)	578 (13.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Unknown	160 (7.9%)	188 (8.3%)	348 (8.1%)	202 (28.2%)	185 (18.0%)	387 (22.2%)
Stage						
Localized	436 (21.4%)	88 (3.9%)	524 (12.2%)	111 (15.5%)	32 (4.5%)	143 (8.2%)
Regional	444 (21.8%)	163 (7.2%)	607 (14.1%)	216 (30.1%)	239 (33.3%)	155 (26.1%)
Distant	1091 (53.6%)	1949 (86.4%)	3040 (70.8%)	244 (34.0%)	635 (88.6%)	879 (50.4%)
Unstaged/unknown	65 (3.2%)	57 (2.5%)	122 (2.8%)	146 (20.4%)	121 (16.9%)	267 (15.3%)
Histology						
Serous	1349 (66.3%)	1943 (86.1%)	3292 (76.7%)	706 (98.5%)	1018 (99.1%)	1724 (98.9%)
Endometrioid	536 (26.3%)	199 (8.8%)	735 (17.1%)	2 (0.3%)	2 (0.2%)	4 (0.2%)
Mucinous	37 (1.8%)	5 (0.2%)	42 (1.0%)	0 (0.0%)	1 (0.1%)	1 (0.1%)
Clear cell	25 (1.2%)	8 (0.4%)	33 (0.8%)	8 (1.1%)	3 (0.3%)	11 (0.6%)
Mixed cell	16 (0.8%)	21 (0.9%)	37 (0.9%)	1 (0.1%)	2 (0.2%)	3 (0.2%)
Other epithelial	73 (3.6%)	81 (3.6%)	154 (3.6%)	0 (0.0%)	1 (0.1%)	1 (0.1%)
Age, mean (SD)	56.8 (11.0)	60.7 (10.8)	58.8 (11.1)	56.7 (10.6)	60.2 (9.7)	58.8 (10.2)

*Grade was based on a 4-tier grading system for Set 1 and a 3-tier grading system for Set 2.

Table 2: Results of single variant EOC overall survival association meta-analysis with meta p<1.0E-4. Analyses are adjusted for age, site, and three PCs.

Chr	rsID (allele)	Gene	Annotation	Set 1 (N=4293)					Set 2 (N=1744)					Meta-Analysis	
				HR (95% CI)	P	MAF	Total number of patients with minor allele	Number of deceased patients with minor allele	HR (95% CI)	P	MAF	Total number of patients with minor allele	Number of deceased patients with minor allele	HR (95% CI)	Meta P
11	rs34335714 (A/G)	CTSW	non-synonymous	5.34 (2.00, 14.25)	0.0008	0.0007	6	4	10.81 (1.52, 76.79)	0.0174	0.0003	1	1	6.15 (2.56, 14.79)	4.91E-5
19	rs8170 (A/G)	*BABAM1*	previous GWAS hit	1.17 (1.09, 1.26)	2.23E-5	0.1966	1446	839	1.14 (1.03, 1.26)	0.0151	0.2124	664	433	1.16 (1.09, 1.23)	1.14E-6
20	rs140079492 (C/T)	PAK7	non-synonymous	3.41 (1.42, 8.19)	0.0062	0.0006	5	5	8.52 (2.74, 26.45)	0.0002	0.0011	4	3	4.80 (2.40, 9.61)	9.32E-6

Table 3: Results of gene-level EOC overall survival meta-analysis, with burden meta p<1.0E-4. CMAF=combined (sum) minor allele frequency across all variants in the gene. Analyses are adjusted for age, site, and three PCs.

Chr	Gene	Set 1 (N=4293)			Set 2 (N=1744)			Burden			SKAT		
		N variants	CMAF	Median MAF	N variants	CMAF	Median MAF	P Set 1	P Set 2	Meta P	P Set 1	P Set 2	Meta P
1	*MTFR1L*	8	0.0032	0.00024	5	0.0026	0.00057	2.21E-03	0.0028	3.13E-05	0.1089	0.0035	0.0037
1	*CD1E*	12	0.5692	0.00047	3	0.3374	0.00057	9.29E-05	0.7255	8.78E-05	0.0002	0.7184	0.0002
3	*HLTF*	17	0.7604	0.00128	7	0.0801	0.00659	1.09E-03	0.0049	1.91E-05	0.0310	0.0538	0.0096
14	*ATG2B*	34	1.7512	0.00058	17	0.4765	0.00057	8.63E-05	0.0030	1.08E-06	0.0790	0.0078	0.0057
19	*PEG3*	38	1.0465	0.00058	18	0.2036	0.00143	3.18E-04	0.0221	1.82E-05	0.0630	0.0856	0.0751

Figure Legends

Figure 1: (A) Manhattan plot of single variant meta-analysis for variants present in both Set 1 and Set 2. –log10(p-values) are plotted by chromosomal location. Survival analysis was adjusted for age, site, and PCs. (B) QQ-plot of single variant meta-analysis; because some variants occur in multiple genes, there are dependencies among tests.

Figure 2: Kaplan-Meier plot based on rs8170 genotype for Set 1 (A) and Set 2 (B).

Figure 3: (A) Manhattan plot of gene-level meta-analysis for genes present in both Set 1 and Set 2. –log10(p-values) are plotted by chromosomal location. Survival analysis was based on the burden test, and adjusted for age, site, and PCs. (B) QQ-plot of gene-level meta-analysis; because some variants occur in multiple genes, there are dependencies among tests.