Genetic Data from Nearly 63,000 Women of European Descent Predicts DNA Methylation Biomarkers and Epithelial Ovarian Cancer Risk

Yang, Yaohua

2019-02-01


http://hdl.handle.net/10138/299088
https://doi.org/10.1158/0008-5472.CAN-18-2726

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
Genetic Data from Nearly 63,000 Women of European Descent Predicts DNA Methylation Biomarkers and Epithelial Ovarian Cancer Risk


1Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee. 2Department of Molecular Physiology & Biophysics, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee. 3Division of Cancer Biostatistics, Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee. 4Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina. 5Department of Epidemiology, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, California. 6Gynecology Unit, Royal Marsden Hospital, London, United Kingdom. 7Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. 8Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain. 9Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, Norway. 10Center for Cancer Biomarkers and Epigenome, University of Bergen, Bergen, Norway. 11Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom. 12Huntsman Cancer Institute, Department of Population Health Sciences, University of Utah, Salt Lake City, Utah. 13Gynecology Research Unit, Haukeland Medical School, Bergen, Norway. 14Sir Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. 15Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 16Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. 17Department of Epidemiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China. 18University of New Mexico Health Sciences Center, University of New Mexico, Albuquerque, New Mexico. 19Department of Cancer Epidemiology and Prevention Research, Alberta Health Services, Calgary, Alberta, Canada. 20Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. 21Obstetrics and Gynecology Epidemiology Center, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts. 22Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, New South Wales, Australia. 23Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, Australia. 24Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom. 25Huntsman Cancer Institute, Department of Population Health Sciences, University of Utah, Salt Lake City, Utah. 26Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. 27Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, United Kingdom. 28Vanderbilt Epidemiology Center, Vanderbilt Genetics Institute, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, Tennessee. 29David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, California. 30Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-MNH, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany. 31The Center for Bioinformatics and Functional Genomics at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California. 32Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia. 33Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. 34Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. 35The Beatson West of Scotland Cancer Centre, Glasgow, United Kingdom. 36Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, Minnesota. 37Samuel Oschin Comprehensive Cancer Institute, Cancer Prevention and Genetics Program, Cedars-Sinai Medical Center, Los Angeles, California.
Abstract

DNA methylation is instrumental for gene regulation. Global changes in the epigenetic landscape have been recognized as a hallmark of cancer. However, the role of DNA methylation in epithelial ovarian cancer (EOC) remains unclear. In this study, high-density genetic and DNA methylation data in white blood cells from the Framingham Heart Study (N = 1,595) were used to build genetic models to predict DNA methylation levels. These prediction models were then applied to the summary statistics of a genome-wide association study (GWAS) of ovarian cancer including 22,406 EOC cases and 40,941 controls to investigate genetically predicted DNA methylation levels in association with EOC risk. Among 62,938 CpG sites investigated, genetically predicted methylation levels at 89 CpG were significantly associated with EOC risk at a Bonferroni-corrected threshold of $P < 7.94 \times 10^{-15}$. Of them, 87 were located at GWAS-identified EOC susceptibility regions and two resided in a genomic region not previously reported to be associated with EOC risk. Integrative analyses of genetic, methylation, and gene expression data identified consistent directions of associations across 12 CpG, five genes, and EOC risk, suggesting that methylation at these 12 CpG may influence EOC risk by regulating expression of these five genes, namely MAPT, HOXB3, ARH1D8, ARHGAP27, and SKAP1. We identified novel DNA methylation markers associated with EOC risk and propose that methylation at multiple CpG may affect EOC risk via regulation of gene expression.

Significance: Identification of novel DNA methylation markers associated with EOC risk suggests that methylation at multiple CpG may affect EOC risk through regulation of gene expression.
Introduction

Ovarian cancer is one of the most deadly cancers among women in the United States (1) and around the world (2). Approximately 90% of ovarian neoplasms are epithelial ovarian cancer (EOC; ref. 1), a heterogeneous disease that can be categorized into five major histotypes (1). Genetic factors have an important impact on EOC etiology. Large-scale genome-wide association studies (GWAS) have identified common risk loci for EOC to date (3). Of these, 27 are specific to the most common histotype, serous EOC (3). However, known loci are estimated to account for only a small proportion (~6.4%) of overall EOC risk (3). In addition, causal genes at most loci and the underlying pathogenic mechanisms are yet to be identified.

In addition to genetic susceptibility, cancer initiation and progression are also influenced by epigenetics (4). The most extensively studied epigenetic marker is DNA methylation, which regulates chromatin structure (5) and gene expression (6). DNA methylation patterns are generally programmed during normal development (7). Abnormal methylation has been observed in multiple malignancies, including EOC (8, 9). Studies have identified multiple DNA methylation markers in tumor tissue samples as prognostic biomarkers for EOC (10, 11). Several studies have also investigated the potential of DNA methylation from white blood cells to be early detection biomarkers for EOC and identified nearly 100 candidate CpGs for EOC risk (12–15). To date, only two CpGs, cg10061138 and cg10636246, were consistently observed across different studies (12–15). The lack of consistent findings may reflect the small sample sizes of prior studies (200–400 cases), an inadequate consideration of potential confounders and reverse causation.

DNA methylation is impacted by both environmental factors and genetic factors (6). High-throughput methylome profiling in both twin and familial studies has shown that methylation levels for a large number of CpGs are heritable (16, 17). Furthermore, several studies (18, 19) have revealed a large number of methylation quantitative trait loci (meQTL) in white blood cells. These results suggest that DNA methylation levels could be partially predicted by genetic variants. Indeed, meQTL single-nucleotide polymorphisms (SNP) appear to predict DNA methylation levels in white blood cells and the predicted methylation levels associated with disease risk (20, 21). However, these studies only used single meQTL SNPs to predict methylation levels for each CpG site. The prediction accuracy is low because meQTL SNPs explain only a small proportion of variance. In this study, we used a novel approach to overcome this limitation by building and validating statistical models to predict methylation levels based on multiple genetic variants in reference datasets. The prediction models were then applied to genetic data from 22,406 cases and 40,941 controls to test the hypothesis that genetically predicted DNA methylation is associated with EOC risk. This approach could overcome the selection bias and reverse causation in conventional epidemiologic studies of DNA methylation and disease because alleles are randomly assigned during gamete formation.

Materials and Methods

Building DNA methylation prediction models using data from the Framingham Heart Study

Genome-wide DNA methylation and genotype data from white blood cell samples from individuals in the Framingham Heart Study (FHS) Offspring Cohort were obtained from dbGaP (accession numbers phs000724 and phs000342, respectively). Detailed descriptions of the FHS Offspring Cohort have been previously reported (22). Genotyping was conducted using the Affymetrix 500K mapping array and imputation was performed with the 1000 Genome Phase 1 (version 3) data as reference. Only SNPs with a minor allele frequency (MAF) of >0.05 and an imputation quality ($R^2$) of >0.80 were used to build prediction models. Genome-wide DNA methylation profiling was generated using the Illumina HumanMethylation450 BeadChip. We used the R package "minfi" (23) to filter low-quality methylation probes, evaluate cell type composition for each sample, and estimate methylation beta-values. Methylation data were then quantile-normalized across samples, rank-normalized to remove potential outliers, and then regressed on covariates including age, sex, cell-type composition, and top ten principal components (PC) to eliminate potential experimental confounders and population structure. Finally, 1,595 unrelated individuals of European descent (883 females and 712 males, mean ± SD of age: 66.3 ± 9.0) with both genetic and DNA methylation data were included in prediction model building.

Using the elastic net method ($\alpha = 0.50$) implemented in the R package "glmnet" (24), we built a statistical model to predict methylation levels for each CpG site using the SNPs within its 2 megabase (Mb) flanking region. For each model, we performed 10-fold cross-validation as internal validation and calculated the squared value of the correlation coefficient between measured and predicted methylation levels, that is, $R_{11F}^2$, to estimate prediction performance.

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

Corresponding Author: Jirong Long, Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN 37203-1728. Phone: 615-343-6741; Fax: 615-343-0719; E-mail: jirong.long@vumc.org.

doi: 10.1158/0008-5472.CAN-18-2726

© 2018 American Association for Cancer Research.

www.aacrjournals.org Cancer Res; 79(3) February 1, 2019 507
Evaluation of model performance using data from the Women’s Health Initiative

Using data from white blood cell samples from 883 independent healthy women of European descent from the Women’s Health Initiative (WHI), we evaluated the performance of the established genetic prediction models. Data from the WHI samples were obtained from dbGaP (accession numbers phs001335, phs000675, and phs00315). Genotyping was conducted using the Illumina HumanMethylation450 BeadChip. The data quality was controlled and imputed using similar criteria and procedures as those described for the FHS data. The Illumina HumanMethylation450 BeadChip was used to profile DNA methylation and the data were then processed using the same pipeline as that for the FHS data. The prediction models established in FHS were applied to the genetic data in WHI to predict methylation levels at each CpG site for each sample. Then, the predicted and measured methylation levels for each CpG site were compared by estimating the squared value of the Pearson correlation coefficient, that is, \( R^2 \).

We used the following criteria to select prediction models for association analyses: (i) a prediction \( R^2_{FHS} \) of \( \geq 0.01 \) (correlation between measured and predicted methylation levels of \( \geq 0.10 \) in the FHS); (ii) a \( R^2_{WHI} \) of \( \geq 0.01 \) in the WHI; and (iii) methylation probes on the HumanMethylation450K BeadChip not overlapping with any SNP included in the dbSNP database (Build 151; ref. 25), considering that SNPs on the probes may have a potential impact on the methylation level estimation (19). In total, models for 63,000 CpGs met these requirements and were included in the downstream association analyses for EOC risk.

Association between genetically predicted DNA methylation and EOC risk

MetaXcan (26) was used to estimate the associations between genetically predicted methylation levels and EOC risk. The methodology of MetaXcan has been described elsewhere (26, 27). Briefly, the following formula was used to evaluate the association Z-score:

\[
Z_m = \frac{\sum_i \text{Model}_i w_m \delta_m \hat{\beta}_m}{\sigma_m se(\hat{\beta}_m)}
\]

In the formula, \( w_m \) represents the weight of SNP \( s \) on the methylation level of the CpG site \( m \), estimated by the prediction model. \( \delta_m \) and \( \sigma_m \) are the evaluated variances of SNP \( s \) and the predicted methylation level at CpG site \( m \), respectively. \( \hat{\beta}_m \) and se(\( \hat{\beta}_m \)) represent the beta coefficient and standard error of SNPs on EOC risk, respectively. For this study, the correlations between predicting SNPs for all CpGs were evaluated using the data from European participants in the 1000 Genomes Project Phase 3.

Beta coefficient \( \beta_m \) and standard error se(\( \hat{\beta}_m \)) for the association between SNP \( s \) and EOC risk were obtained from the Ovarian Cancer Association Consortium (OCAC), which includes 22,406 EOC cases and 40,941 controls of European ancestry (3). Details of this consortium have been described elsewhere (3). For patients with EOC, some may have had neo-adj chemotherapy before surgery. They were not included in subtype analyses but included in the analyses for overall EOC risk (3). Cases were classified as one of five histotypes: high-grade serous \( (n = 13,037) \), endometrioid \( (n = 2,810) \), mucinous invasive \( (n = 1,417) \), clear cell \( (n = 1,366) \), or low-grade serous \( (n = 1,012) \). In addition, there were 2,764 EOC cases that could not be categorized into any histotypes. Genotyping was conducted using OncoArray and other GWAS arrays, followed by imputation, with the 1000 Genomes Project Phase 3 as reference. Association analyses were conducted within each dataset (different GWAS arrays) and the results were combined by a fixed-effect inverse-variance meta-analysis. Among the 751,157 SNPs included in the prediction models for 63,000 CpGs, summary statistics for associations between 751,031 (99.98%) of these SNPs and EOC risk were available from the OCAC. A total of 62,938 CpGs, corresponding to these 751,031 SNPs, were included in the final analyses. This study was approved by the OCAC Data Access Coordination Committee.

For risk analyses in OCAC, we used a Bonferroni-corrected threshold of \( P < 7.94 \times 10^{-7} \) (0.05/62,938) for statistical significance in assessing the association between each of the 62,938 CpGs and EOC risk. Associations of predicted methylation and EOC risk identified in the OCAC data were further evaluated using the summary statistics of two GWAS studies of ovarian cancer in the UK Biobank (28). However, the sample size of the EOC cases is very small, with only 440 histologically diagnosed and 579 self-reported ovarian cancer cases among nearly 337,000 unrelated individuals of European descent. GWAS analyses were conducted using a linear regression model. The summary statistics data are available at https://sites.google.com/broadinstitute.org/ukbbwes/results/home.

We estimated whether the identified associations of predicted methylation with EOC risk were independent of GWAS-identified EOC susceptibility variants. For each SNP included in the prediction model, we used GCTA-COJO (29) to evaluate the \( \beta_m \) with EOC risk after adjusting for the GWAS-identified variants for EOC. Then, we reconducted the MetaXcan analyses to investigate the associations of the predicted methylation levels with EOC risk conditioning on the GWAS-identified EOC risk variants. We also performed stratification analyses by six EOC histotypes and estimated the heterogeneity across histotype groups by using Cochran Q test.

Functional annotation of methylation markers

Using ANNOVAR (30), all 62,938 investigated CpGs were classified into 11 functional categories: upstream, transcription start site upstream 1,500 bp (TSS1500), TSS200, 5’-untranslated region (UTR), exonic, intronic, 3′-UTR, downstream, intergenic, noncoding RNA (ncRNA) exonic and ncRNA intronic.

Correlation analyses of DNA methylation with gene expression in white blood cells

For those 89 CpGs with predicted methylation levels associated with EOC risk, we investigated those methylation levels in relation to the expression levels of genes flanking these CpGs. Individual-level DNA methylation and gene expression data of white blood cell samples from the FHS Offspring Cohort were accessed from dbGaP (accession numbers phs000724 and phs000363). The details of the Offspring Cohort of the FHS, the DNA methylation data and gene expression data have been described previously (22, 31). In total, 1,367 unrelated participants with both methylation and gene expression data were included in correlation analyses. A threshold of \( P < 0.05 \) was used to determine a nominally significant correlation between methylation level and gene expression level. In addition, using data from the FHS, we...
investigated whether the methylation of those 89 EOC-associated CpGs could regulate the expression of 19 homologous recombination (HR) genes (32, 33).

**Association analyses of genetically predicted gene expression with EOC risk**

For genes with expression levels nominally correlated with methylation levels of CpGs that were associated with EOC, we investigated whether genetically predicted gene expression levels were associated with EOC risk following methods described elsewhere (27). Briefly, genome-wide genetic and gene expression data from 6,124 different tissue samples (donated by 369 participants of European ancestry) included in the Genotype-Tissue Expression (GTEx) release 6 (34) were used to build genetic models for gene expression prediction by following the elastic net method (27). The models were then applied to the OCAC data to estimate the associations between genetically predicted gene expression levels and EOC risk by using MetaXcan (26). We used Bonferroni correction to declare statistically significant associations.

**Consistent directions of associations across methylation, gene expression, and EOC risk**

To infer potential mechanisms underlying the identified associations between DNA methylation and EOC risk, we conducted integrative analyses of the association results between predicted CpG methylation and EOC risk, correlations between CpG methylation and gene expression, and associations between gene expression and EOC risk. First, we examined whether the association directions among DNA methylation, gene expression and EOC risk were consistent. Then, we evaluated whether genetically predicted methylation might mediate associations between gene expression and EOC risk. Briefly, for each gene we used GCTA-COJO (35) to generate modified summary statistics of associations between SNPs in its expression prediction models and EOC risk after adjusting for SNPs included in the methylation prediction model of its corresponding CpG site. Finally, the prediction models of each gene were applied to the updated summary statistics using MetaXcan (26) to estimate the association between genetically predicted gene expression and EOC risk conditioning on the effects of the genetically predicted methylation level at each corresponding CpG site.

**Results**

**DNA methylation prediction models**

Figure 1 presents the overall workflow of this study. Data from the FHS Offspring Cohort were used to create methylation prediction models for 223,959 CpGs. Of these, 81,361 showed a prediction performance ($R^2_{FHS}$) of >0.01, representing at least a 10% correlation between predicted and measured methylation levels. For these 81,361 CpGs, the number of SNPs in prediction models ranged from 1 to 276, with a median of 25. Applying these 81,361 models to genetic data from the WHI, 70,269 (86.4%) models showed a correlation coefficient between predicted and measured methylation levels ($R^2_{WHI}$) of >10%. Among these 70,269 CpGs, methylation probes of 7,269 on the HumanMethylation450 BeadChip overlapped with SNPs, which may have affected the estimation of the methylation levels (19). Hence, these CpGs were excluded. The remaining 63,000 CpGs were included in the downstream analyses.

**Associations of genetically predicted DNA methylation with EOC risk**

The prediction models were applied to the data from a GWAS of 22,406 EOC cases and 40,941 controls included in OCAC. Summary statistics of associations between 751,031 of the 751,157 SNPs, corresponding to 62,938 of the 63,000 CpGs, and EOC risk were available in OCAC. For these 62,938 CpGs, a high correlation of prediction performance between models based on FHS
(Refs2) and WHI (Refs2) data was observed, with a Pearson correlation coefficient of 0.95. This indicates that for each of these CpGs, a same set of predicting SNPs could predict a very similar methylation level, using either FHS or WHI data.

For most of these 62,938 CpGs, a large majority of predicting SNPs were available in OCAC (e.g., for 94% of the investigated CpGs ≥95% of the SNPs in prediction models were available in OCAC). Supplementary Figure S1 is the Manhattan plot presenting the associations between genetically predicted methylation levels and EOC risk. Among the 62,938 CpGs, 89 were significantly associated with EOC risk at a Bonferroni-corrected threshold of \( P < 7.94 \times 10^{-7} \) (Tables 1 and 2; Supplementary Table S1).

Among these 89 CpGs, a higher predicted methylation level was associated with an increased risk of EOC at 48 CpGs, and with a decreased EOC risk in the other 41 CpGs. This indicates that the methylation levels were predicted to be higher for 48 CpGs and lower for 41 CpGs among EOC cases than among controls. For these 89 CpGs, we also rebuilt the prediction models only using data from females \( (N = 833) \) in FHS. A very high correlation was observed, with a Pearson correlation coefficient of 0.99, between the prediction performance \( R^2 \) values, based on data of all FHS participants \( (N = 1,595) \) and those based on females only data \( (N = 833) \). In the UK Biobank data, consistent associations were observed for 23 CpGs, including 12 at \( P < 0.05 \), and 11 additional CpGs at \( P < 0.10 \) (Supplementary Table S2). This relatively low replication rate is not unexpected, considering the very limited statistical power of the UK Biobank data because of a very small number of cases \( (N = 400–600) \).

Among the 89 CpGs that were associated with EOC, two reside in a genomic region on chromosome 7 that has not yet been reported for EOC risk (500 Kb away from any GWAS-identified EOC susceptibility variants; Table 1). Given that there are no risk variants identified by previous GWAS on this chromosome, associations with EOC risk conditioning on proximally located risk variants could not be conducted. Among the remaining 87 CpGs located in nine previously identified EOC risk loci, no associations remained significant after an adjustment for all risk SNPs in the corresponding loci. This suggests that the associations of these 87 CpGs with EOC risk were all driven by known EOC risk SNPs in these loci (Table 2, Supplementary Table S1).

Stratification analyses by EOC histotypes revealed that all 89 CpGs were associated with both serous ovarian cancer and high-grade serous ovarian cancer. Fewer CpGs were associated with the other histotypes, including endometrioid ovarian cancer (cg25137403, cg14454907, and cg25708328), mucinous ovarian cancer (cg25137403, cg14454907, cg10086659, and cg25708328), and low-grade serous ovarian cancer (cg01572694; Supplementary Tables S3–S4). Fourteen of these 89 CpGs showed more significant associations with the serous and the high-grade serous ovarian cancers than with other histotypes, with a heterogeneity test \( P < 5.62 \times 10^{-4} \), a Bonferroni-corrected threshold (0.05/89; Supplementary Table S3). Among these 89 CpGs, a significant correlation of methylation and gene expression was identified for 91 CpG-HP gene pairs, including 22 CpGs and 11 HR genes, at a Bonferroni-corrected threshold of \( P < 2.96 \times 10^{-5} \) (0.05/1,691; Supplementary Table S5). Interestingly, methylation levels of three CpGs, that is, cg13568213 (9q34.2), cg10900703 (10p12.31), and cg23659289 (17q21.31) showed a strong correlation with the expression level of the ATM gene.

**DNA methylation affecting EOC risk through regulating expression of a neighbor gene**

For those 89 CpGs with predicted methylation levels associated with EOC risk, we conducted correlation analyses with gene expressions for 63 pairs of CpG-gene, including 58 CpGs with 21 flanking genes that were annotated by ANNOVAR (30). Nominally significant correlations were observed for 26 CpG-gene pairs, including 26 CpGs and 12 genes, at \( P < 0.05 \) (Table 3; Supplementary Table S6). Among them, the most significant correlation was observed between the increased methylation at the CpG cg19139618, located in the promoter region of the SKAPI gene, and the expression level of SKAPI, with a \( P \) value of \( 2.98 \times 10^{-15} \) (Table 3). In addition, increased methylation levels at two CpGs, cg10900703 and cg04231319, located in the introns of the MLLT10 gene, were significantly correlated with an increased expression of MLLT10, with \( P \) values of \( 2.79 \times 10^{-11} \) and \( 1.36 \times 10^{-5} \), respectively. For the two CpGs located in a putative novel locus, a higher methylation level for one of them, cg03634833, was correlated with a lower expression of the ADAP1 gene in this locus, with a \( P \) value of \( 2.99 \times 10^{-4} \) (Supplementary Table S6). As expected, methylation levels at CpGs located in promoter regions (TSS1500 and TSS200) were more likely to be negatively correlated with expressions of proximal genes. Nearly all CpGs located in downstream or in 3’UTR showed a negative

---

**Table 1. Two novel methylation-EOC associations for two CpGs located at a genomic region not yet reported for EOC risk**

<table>
<thead>
<tr>
<th>CpG</th>
<th>Chr</th>
<th>Position</th>
<th>Closest gene</th>
<th>Classification</th>
<th>( R_{max} ) (^a)</th>
<th>Histotype</th>
<th>( Z ) score</th>
<th>OR (95% CI) (^b)</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg18139273</td>
<td>7</td>
<td>962,582</td>
<td>ADAP1</td>
<td>Intrinsic</td>
<td>0.01</td>
<td>Overall</td>
<td>–4.95</td>
<td>0.51 (0.39–0.66)</td>
<td>7.25 ( \times 10^{-7} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Serous</td>
<td>–4.87</td>
<td>0.46 (0.34–0.63)</td>
<td>1.13 ( \times 10^{-6} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High-grade serous</td>
<td>–4.83</td>
<td>0.46 (0.33–0.63)</td>
<td>1.39 ( \times 10^{-6} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Endometrioid</td>
<td>–1.78</td>
<td>0.59 (0.35–1.06)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mucinous</td>
<td>–0.99</td>
<td>0.67 (0.30–1.49)</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clear cell</td>
<td>–1.87</td>
<td>0.46 (0.21–1.04)</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low-grade serous</td>
<td>–0.97</td>
<td>0.62 (0.24–1.63)</td>
<td>0.33</td>
</tr>
<tr>
<td>cg03634833</td>
<td>7</td>
<td>965,534</td>
<td>ADAP1</td>
<td>Intrinsic</td>
<td>0.09</td>
<td>Overall</td>
<td>–5.00</td>
<td>0.84 (0.79–0.90)</td>
<td>5.81 ( \times 10^{-7} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Serous</td>
<td>–4.85</td>
<td>0.83 (0.77–0.89)</td>
<td>1.21 ( \times 10^{-6} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High-grade serous</td>
<td>–4.85</td>
<td>0.82 (0.76–0.89)</td>
<td>1.25 ( \times 10^{-6} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Endometrioid</td>
<td>–2.21</td>
<td>0.85 (0.71–0.98)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mucinous</td>
<td>–1.40</td>
<td>0.87 (0.71–1.06)</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clear cell</td>
<td>–1.76</td>
<td>0.84 (0.69–1.02)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low-grade serous</td>
<td>–0.87</td>
<td>0.91 (0.73–1.13)</td>
<td>0.39</td>
</tr>
</tbody>
</table>

**Abbreviation:** CI, confidence interval.

\(^a\)Correlation between predicted and measured methylation levels.

\(^b\)OR per SD increase in genetically predicted methylation level.

\(^c\)Including high-grade serous and low-grade serous ovarian cancers.
DNA Methylation Biomarkers and Epithelial Ovarian Cancer

For the 12 genes with expression levels correlated with DNA methylation, expression prediction models were built for seven, with a prediction performance ($R^2$) of $\geq 0.01$, using GTEx data. Applying these seven models to the OCAC data, genetically predicted expression levels of three genes, namely MAPT, HOXB3, and ABHD8, were significantly associated with EOC risk after Bonferroni correction (Table 4). At 17q21.31 and 17q21.32, higher predicted expression levels of MAPT and HOXB3 were associated with a decreased EOC risk, with $P$-values of $3.74 \times 10^{-4}$ and $2.00 \times 10^{-5}$, respectively. After adjusting for established EOC risk SNPs, the associations between these two genes and EOC risk disappeared. At 19p13.11, an increased predicted expression level for ABHD8 was associated with an increased EOC risk, with a $P$-value of $9.93 \times 10^{-6}$. Conditioning on the EOC risk SNP in this locus, the association disappeared as well (Table 4). Of the five genes without prediction models, two were previously reported to be associated with EOC susceptibility, including SKAPI (36) and ARHGAP27 (37).

We integrated the results for the association between DNA methylation and EOC risk, the correlation between DNA methylation and gene expression, and the association between gene expression and EOC risk. We identified consistent directions of associations across seven CpGs, including cg18878992, cg00480298, cg07368061, cg01572694, cg14285150, cg24672833, and cg17941109, three genes, including MAPT, HOXB3 and ABHD8, and EOC risk (Table 5). The mechanism potentially underlying the associations of methylation at these seven CpGs and EOC risk may be their regulatory function on expression of these three genes. Among them, increased methylation at the CpG site cg14285150 was associated with an increased HOXB3 expression ($P = 8.44 \times 10^{-5}$) and decreased EOC risk ($P = 5.53 \times 10^{-4}$). As expected, an increased expression of HOXB3 was associated with a decreased EOC risk ($P = 2.00 \times 10^{-7}$). Conditioning on SNPs included in the methylation prediction model for cg14285150, the association of HOXB3 expression and EOC risk disappeared ($P = 0.51$; Table 5).

Expression prediction models could not be built for SKAPI at 17q21.32 and ARHGAP27 at 17q21.31 in this study. Hence, these two genes could not be investigated in association with EOC risk. However, higher expression levels of these two genes have been previously reported to be associated with an increased risk of EOC (36, 37). This is expected, based on the association results of DNA methylation with EOC risk and DNA methylation with gene expression (Table 5). For example, a higher methylation at cg19139618 was associated with a lower expression of SKAPI ($P = 2.98 \times 10^{-3}$) and lower EOC risk ($P = 7.08 \times 10^{-2}$). Hence, the potential mechanism underlying the association between cg19139618 and EOC risk may be the downregulation effects on SKAPI expression (Table 5).

**Discussion**

In this large study, we identified 89 CpGs that were significantly associated with EOC risk, including two CpGs located in a novel genomic region that have not yet been reported as a susceptibility locus for EOC. Integrating genetic, methylation, and gene expression data suggested that methylation at 12 of 89 CpGs may exert their impacts on EOC risk through regulating the expression of regulatory effect on expression of neighbor genes. For CpGs residing in intronic regions, both positive and negative correlations were observed.
five genes. These results provide new insights into the regulatory pathways that connect genetics, epigenetics, gene expression, and EOC risk.

We identified two methylation markers, cg18139273 and cg03634833, located at 7p22.3, a novel genomic region that had not been reported as a risk locus for EOC. Both CpGs reside in the third intron of the first transcript of the ADAP1 gene, which encodes an ADP-ribosylation factor GTPase-activating protein (ArfGAP) with dual PH domains 1. ADAP1 functions as a scaffolding protein in several signal transduction pathways. It is highly expressed in neurons, where it has roles in neuronal differentiation and neurodegeneration (38). This gene has also been reported to be involved in microfibril function (39), and is a target of the ErbB4 transcription factor in mammary epithelial cells (40). In this study, we found that a higher methylation level at cg03634833 was significantly correlated with a lower ADAP1 expression, which was associated with a nonsignificantly decreased EOC risk. Thus, methylation at cg03634833 might be associated with EOC risk through a regulatory function on ADAP1 expression, or through other unidentified mechanisms.

Integrating the results of the association between DNA methylation and EOC risk, the correlation between DNA methylation and gene expression, and the association between gene expression and EOC risk, we observed consistent directions of associations across 12 CpGs, five genes, and EOC risk. For the MAPT gene (17q21.31), an increased methylation at two CpGs located in its exons, cg18878992 and cg00480298, were associated with a decreased MAPT expression and increased EOC risk. For the other CpG site, cg07368061, located at the first intron of MAPT, its increased methylation was associated with a higher MAPT expression and lower EOC risk. As expected, an increased MAPT expression was associated with decreased EOC risk. The MAPT gene has been linked to multiple neurodegenerative disorders, including progressive supranuclear palsy (41), Parkinson’s disease (42, 43), and Alzheimer’s disease (42). In addition, a higher expression of a MAPT protein isoform (<70 kDa) was correlated with a lower sensitivity to taxanes in breast cancer cells (44). Methylation of the miRNA miR-34c-5p was shown to regulate the MAPT expression, which was related to paclitaxel resistance in gastric cancer cells (45).

Increased methylation of three CpGs in the first intron of the HOXB3 gene (17q21.32), cg01572694, cg14285150, and cg24672833, were associated with an increased expression of HOXB3 and decreased EOC risk. As expected, an increased HOXB3 expression was associated with decreased EOC risk. A previous study reported that the expression of HOXB3 was upregulated in EOC cell lines compared with normal samples (46). However, this study only included 5 patients and the results have not been replicated by an independent study. On the other side, we investigated the genetically predicted methylation levels in DNA from white blood cells, but not in ovary or fallopian tube epithelial cells. It is possible that the correlation between methylation levels of these CpGs and HOXB3 expression are different in ovary epithelial cells and white blood cells. For example, in the 5’UTR of HOXB3, a higher methylation at the CpG cg12910797 was significantly associated with an increased EOC risk. The increased methylation of this CpG was not correlated with the expression of HOXB3 in white blood cells samples from the FHS (Spearman correlation coefficient $r = -0.02; P = 0.43$). Higher methylation of this CpG was significantly correlated with a decreased HOXB3 expression in ovarian serous cystadenocarcinoma samples from the Cancer Genome Atlas (TCGA) (Spearman correlation coefficient $r = -0.27; P = 2.01 \times 10^{-6}$; http://gdac.broadinstitute.org/runs/analyses__2016_01_28/reports/cancer/ OV-TP/Correlate_Methylation_vs_miRNA/nozzle.html).

The higher methylation of the CpG site cg17941109, located at the second intron of the ABHD8 gene, was associated with a lower ABHD8 expression and a lower EOC risk. This is consistent with the results of two recent studies that showed that a higher expression level of this gene was associated with an increased risk of EOC (47, 48). This gene is located at 19p13.11, a susceptibility locus for both ovarian and breast cancers. Interestingly, in our unpublished data, the increased genetically predicted methylation level at cg17941109 was associated with decreased breast cancer risk, and the genetically predicted expression of ABHD8 was associated with an increased breast cancer risk. Increasing evidence also suggests that this protein family (ABHD) has a physiologic significance in metabolism and disease (49).

For the ARHGAP27 gene, increased methylation of two CpGs in the promoter region, cg16281322 and cg25708777, and one CpG in the 3’-UTR, cg07067577, were associated with lower expression

<table>
<thead>
<tr>
<th>Table 3.</th>
<th>Selected correlations between methylation levels at 26 CpGs and expression levels of 12 genes; data from the FHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpG</td>
<td>Chr Position</td>
</tr>
<tr>
<td>cg02517403</td>
<td>2</td>
</tr>
<tr>
<td>cg2221092</td>
<td>7</td>
</tr>
<tr>
<td>cg03634833</td>
<td>7</td>
</tr>
<tr>
<td>cg14653977</td>
<td>9</td>
</tr>
<tr>
<td>cg24367699</td>
<td>9</td>
</tr>
<tr>
<td>cg10900703</td>
<td>10</td>
</tr>
<tr>
<td>cg23659289</td>
<td>17</td>
</tr>
<tr>
<td>cg07368061</td>
<td>17</td>
</tr>
<tr>
<td>cg19396181</td>
<td>17</td>
</tr>
<tr>
<td>cg14285150</td>
<td>17</td>
</tr>
<tr>
<td>cg22312020</td>
<td>17</td>
</tr>
<tr>
<td>cg79141109</td>
<td>19</td>
</tr>
</tbody>
</table>

*Selected from correlations between 26 CpGs and 12 genes. For each gene, only the most significantly correlated CpG is presented. Complete list of results for all CpG–EOC associations is available in Supplementary Table S6.

| ABHD8     | 19p13.11 |
| GBGT1     | 10p13.12 |
| HOXB3     | 17q21.31 |
| HOXB8     | 17q21.31 |
| HOXD4     | 17q21.31 |
| HOXD8     | 10p13.12 |
| MAPT      | 17q21.31 |
| MLLT10    | 12q24.13 |
| ADAP1     | 17q21.31 |
| ABO       | 17q21.31 |
| SKAPI     | 17q21.31 |
| ARHGAP27  | 17q21.31 |

Table 3. Selected correlations between methylation levels at 26 CpGs and expression levels of 12 genes; data from the FHS

<table>
<thead>
<tr>
<th>Region</th>
<th>Gene Type</th>
<th>Z score</th>
<th>P</th>
<th>Padj</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>17q21.31</td>
<td>MAPT</td>
<td>Protein</td>
<td>-5.96</td>
<td>3.74 \times 10^{-5}</td>
<td>0.40</td>
</tr>
<tr>
<td>17q21.31</td>
<td>HOXB3</td>
<td>Protein</td>
<td>-5.20</td>
<td>2.00 \times 10^{-7}</td>
<td>0.71</td>
</tr>
<tr>
<td>19p13.11</td>
<td>ABHD8</td>
<td>Protein</td>
<td>4.42</td>
<td>9.93 \times 10^{-6}</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*Adjusted for the EOC risk SNPs in the corresponding locus.

*Correlation between predicted and measured gene expression levels.
level of ARHGAP27 and lower EOC risk. For the SKAP1 gene, a higher methylation at the CpG cg02957270, located at the promoter region, was associated with a higher expression level and increased EOC risk. Increased methylation of the other intronic CpG, cg19139618, was associated with a lower SKAP1 expression and a decreased EOC risk. In this study, the associations of expression levels of these two genes and EOC risk could not be investigated because the prediction models for them could not be built. However, two large GWAS studies have identified these two genes as EOC susceptibility genes with solid experimental evidence (36, 37). Differential expression analyses showed a significantly higher expression of ARHGAP27 in ovarian cancer than in normal cells (37). It is suggested that the ARHGAP27 gene may play a role in carcinogenesis through the dysregulation of Rho/Rac/Cdc42-like GTPases (50). The expression of SKAP1 was significantly greater in ovarian cancer cells when compared with primary human ovarian surface epithelial cells (36). Our study is the first to suggest that these two genes may be associated with EOC risk through methylation regulation.

Several epidemiologic studies have investigated the associations of CpG methylation and EOC risk in white blood cells and tumor tissue samples (12–15). Approximately 100 CpGs have been identified to be associated with EOC risk. However, only two CpGs, cg10061138 and cg10636246, showed consistent association directions in two or more studies. In this study, prediction models could not be built for these two CpGs; hence, neither could be investigated in association with EOC risk. Among the remaining 98 reported CpGs, reliable prediction models were built for only 20 of them and only two, cg19399532 and cg21870884, could be replicated at \( P < 0.10 \), with the same association directions as previously reported. Such a low replication rate is not unexpected because of several potential limitations in traditional epidemiologic studies, which include possible false associations because of small sample size, lack of validation in other studies, potential confounders, and reverse causation.

The methodology of this study is similar to that of transcriptome-wide association studies (TWAS), in which gene expression prediction models are established and applied to GWAS data to investigate genetically predicted gene expression in association with various diseases and traits. Of the five genes identified in this study, the expression levels of two, HOXB3 and ABHD8, were significantly associated with EOC risk at the Bonferroni-corrected threshold (\( P < 2.2 \times 10^{-8} \)) in our previous TWAS study for EOC (51). The MAPT gene showed an association with EOC at \( P = 3.74 \times 10^{-4} \) in the TWAS; however, the association did not reach the Bonferroni-corrected threshold. For ARHGAP27 and SKAP1, gene expression prediction models could not be built, and they were not investigated in the TWAS. Expression levels for these two genes were reported to be associated with EOC (36, 37). Some genes identified in TWAS were not tested in this study because the methylation prediction models could not be built for CpGs flanking them. In addition, except DNA methylation, there are other biological processes that regulate gene expression. The regulation of DNA methylation on gene expression differs according to the locations of the CpGs. Therefore, integrating the results of methylation and gene expression analyses may help to understand the biological basis for EOC.

It would be ideal to build methylation prediction models using data from normal ovary or fallopian tube epithelial cells, but it is almost impossible to collect tissue samples from a large population of healthy women. However, as demonstrated by multiple studies, the large majority of the meQTLs identified in white blood cells were consistently detected across different tissue types (26, 52, 53). These results indicate that the genetically determined methylation at many CpGs are predictable and consistent among different tissues. Hence, it is reasonable to build methylation prediction of DNA methylation in association with EOC. It would be ideal to validate the findings in this study by directly measuring methylation levels in prediagnosis blood samples in prospective studies to overcome reverse causation; however, the majority of the samples included in this study were collected after cancer diagnosis. It is possible that DNA methylation regulation on gene expression differs across tissues. In this study, data in white blood cell samples were used, which is another limitation. In the association analysis of predicted gene expression with EOC risk, the models were built using data from a limited sample size of GTEx. Thus, the number of genes evaluated in our study was small. More consistent associations across methylation, gene expression, and EOC risk could be identified with a larger sample size to build gene expression prediction models.

Strengths of this study include the large number of samples in the reference dataset used in model building and that the model performance was evaluated in an independent dataset. Using genetic variants as study instruments, we can effectively overcome

<table>
<thead>
<tr>
<th>CpG</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>Classification</th>
<th>CpG vs. EOC risk</th>
<th>CpG vs. Gex</th>
<th>Gex vs. EOC risk</th>
<th>Adjusted* Gex vs. EOC risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg0878992</td>
<td>17</td>
<td>43,974,344</td>
<td>MAPT</td>
<td>5’UTR</td>
<td>+ 8.65 x 10^{-15}</td>
<td>+ 2.64 x 10^{-1}</td>
<td>- 3.74 x 10^{-4}</td>
<td>- 0.48</td>
</tr>
<tr>
<td>cg00428928</td>
<td>17</td>
<td>40,106,390</td>
<td>Exonic</td>
<td>- 6.39 x 10^{-9}</td>
<td>- 3.98 x 10^{-3}</td>
<td>- 3.74 x 10^{-4}</td>
<td>- 0.65</td>
<td></td>
</tr>
<tr>
<td>cg07360601</td>
<td>17</td>
<td>44,090,862</td>
<td>MAPT</td>
<td>Intronic</td>
<td>- 4.26 x 10^{-12}</td>
<td>+ 2.02 x 10^{-2}</td>
<td>- 3.74 x 10^{-4}</td>
<td>- 1.00</td>
</tr>
<tr>
<td>cg00715709</td>
<td>17</td>
<td>46,575,555</td>
<td>HOXB3</td>
<td>Intronic</td>
<td>+ 5.52 x 10^{-9}</td>
<td>+ 7.49 x 10^{-3}</td>
<td>- 2.00 x 10^{-7}</td>
<td>- 0.82</td>
</tr>
<tr>
<td>cg14285150</td>
<td>17</td>
<td>46,659,099</td>
<td>HOXB3</td>
<td>Intronic</td>
<td>- 5.53 x 10^{-8}</td>
<td>+ 8.44 x 10^{-5}</td>
<td>- 2.00 x 10^{-7}</td>
<td>- 0.51</td>
</tr>
<tr>
<td>cg24672833</td>
<td>17</td>
<td>46,659,318</td>
<td>HOXB3</td>
<td>Intronic</td>
<td>- 9.00 x 10^{-9}</td>
<td>+ 5.51 x 10^{-2}</td>
<td>- 2.00 x 10^{-7}</td>
<td>- 0.41</td>
</tr>
<tr>
<td>cg79412610</td>
<td>17</td>
<td>47,007,199</td>
<td>ABHD8</td>
<td>Intronic</td>
<td>+ 2.88 x 10^{-9}</td>
<td>+ 0.03</td>
<td>- 9.93 x 10^{-6}</td>
<td>- 0.57</td>
</tr>
<tr>
<td>cg1939618</td>
<td>17</td>
<td>46,504,791</td>
<td>SKAP1</td>
<td>Intronic</td>
<td>+ 7.08 x 10^{-10}</td>
<td>- 2.98 x 10^{-15}</td>
<td>+</td>
<td>NA^b</td>
</tr>
<tr>
<td>cg02957270</td>
<td>17</td>
<td>46,508,097</td>
<td>SKAP1</td>
<td>TSS5000</td>
<td>+ 4.40 x 10^{-12}</td>
<td>- 0.01</td>
<td>-</td>
<td>NA^b</td>
</tr>
<tr>
<td>cg70675770</td>
<td>17</td>
<td>45,506,829</td>
<td>ARHGAP27</td>
<td>3’UTR</td>
<td>+ 6.86 x 10^{-14}</td>
<td>+ 1.20 x 10^{-2}</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>cg16821322</td>
<td>17</td>
<td>43,510,478</td>
<td>ARHGAP27</td>
<td>TSS2000</td>
<td>+ 6.82 x 10^{-13}</td>
<td>+ 1.14 x 10^{-9}</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>cg25078777</td>
<td>17</td>
<td>43,510,841</td>
<td>ARHGAP27</td>
<td>TSS5000</td>
<td>+ 4.61 x 10^{-13}</td>
<td>- 4.11 x 10^{-8}</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Dir, direction of association/correlation; Gex, gene expression.

*aAdjusting for all the predicting SNPs included in prediction models of corresponding CpGs.

*SKAP1 and ARHGAP27 are previously identified EOC-susceptibility genes.
many limitations commonly encountered in conventional epidemiologic studies. In addition, this is the largest study of DNA methylation with EOC risk and a very stringent criterion was used, providing high statistical power to identify reliable associations between genetically predicted methylation and EOC risk. Finally, the integrative analyses of genetic, DNA methylation, and gene expression data led to the identification of consistent evidence to support the hypothesis that DNA methylation could impact EOC risk through regulating gene expression.

In summary, in the largest study conducted to date that investigates DNA methylation in association with EOC risk, we identified multiple CpGs that were significantly associated with EOC risk and proposed that several CpGs may affect EOC risk through regulating expression of five genes. Our study demonstrates the feasibility of integrating multi-omics data to identify novel biomarkers for EOC risk and brings new insight into the etiology of this malignancy.

Disclosure of Potential Conflicts of Interest

S. Banerjee is a consultant/advisory board member for AstraZeneca, Tesaro, Clovis, Roche, Gamavams, Merck, Seattle Genetics, and Pharmamar and has provided expert testimony for AstraZeneca, Tesaro, and Roche. J.D. Brenton reports receiving a commercial research grant from Aprea and has ownership interest (including stock, patents, etc.) in Inivata Ltd. A. DeFazio reports receiving other commercial research support from AstraZeneca. P.A. Fasching reports receiving a commercial research grant from Novartis to Institution, is a consultant/advisory board member for Novartis, Roche, Celgene, Pfizer, DataSankeyo, Teva, and Puma. B.Y. Karlan is a consultant/advisory board member for Invitae Corporation. I.A. McNeish is a consultant/advisory board member for Clong Oncology, AstraZeneca, Tesaro, and Takeda. U. Menon has ownership interest (including stock, patents, etc.) in Abcodia Pvt Ltd. Y.L. Woo has received speaker bureau honoraria from Merck Sharp and Dohme and is a consultant/advisory board member for Merck Sharp and Dohme. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions


Development of methodology: Y. Yang, L. Wu, D.G. Huntsman, K.B. Moyaich


Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Yang, L. Wu, Y. Lu, C. A. Beigbeder-Fadel, B. Li, F. Ye, J. Dennis, R.T. Fortner, S.A. Gayther, B.Y. Karlan, D. Lambrecht, T. May, K.B. Moyaich, H. Olsson, S.K. Park, J.P. Tyrer, N. Wentzensen, E. White, A. Wolk, D. Yannakakos


Other (providing data): R.B. Ness

Acknowledgments

We thank Jing He and Marshall Younger of Vanderbilt Epidemiology Center for their help with this study. The authors wish to thank all of the individuals who took part in the study, and all of the researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. Data for the FHS Offspring Cohort were obtained from dbGaP (accession numbers phs000724, phs003432, and phs003636). Data for the WHI were obtained from dbGaP (accession numbers phs001335, phs000675 and phs003153). The data analyses were conducted using the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.

Acknowledgments for individual studies: AOV: We thank Jennifer Kozial, Mie Konno, Michelle Darago, Fay Chambers and the Tom Baker Cancer Centre Translational Laboratories, AIS: The AOCS also acknowledges the cooperation of the participating institutions in Australia and acknowledges the contribution of the study nurses, research assistants and all clinical and scientific collaborators to the study. The complete AOSCS Study Group can be found at www.aosct.org. We would like to thank all of the women who participated in these research programs, BEL: We would like to thank Gilian Feurman, Thomas Van Brussel, Annick Van den Broeck and Joke De Roover for technical assistance. BGS: The BGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR). ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. We thank the study staff, study participants, doctors, nurses, health care providers, and health information sources who have contributed to the study. BUU: The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center’s BioVU, which is supported by institutional funding, the 1510RR025141-01 instrumentation award, and by the Vanderbilt CTSA grant UL1TR000445 from NCATS/NIH; CAM: This work was supported by the National Cancer Institute for Health Research Centre. We thank the study staff, study participants, doctors, nurses, health care providers, and health information sources who have contributed to the study; CHA: Innovative Research Team in University (PCSRT) in China (IRT1076); CHN: We thank all members of Department of Obstetrics and Gynecology, Hebei Medical University, Fourth Hospital and Department of Molecular Biology, Hebei Medical University, Fourth Hospital; CIN: Cytogenetic Cancer Center of Excellence (W61XWH11-2-0131); COBAS: the cooperation of the 32 Connecticut hospitals, including Stamford Hospital, in allowing patient access, is gratefully acknowledged. This study was approved by the State of Connecticut Department of Public Health Human Investigation Committee. Certain data used in this study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. The authors assume full responsibility for analyses and interpretation of these data; DKE: OCBR; EPC: We thank all members and investigators of the Rotterdam Ovarian Cancer Study. Dutch Cancer Society (EMC 2014-6499); GER: The German Ovarian Cancer Study (GER) thank Uruna Elber for competent technical assistance; HOC: The study was supported by the Helsinki University Research Fund, JGO: ISPS KAKENHI grant; KRA: This study (Ko-EVE) was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) and the National R&D Program Ministry of Health & Welfare, Republic of Korea (HI116C1127; D2001000). LUN: ECC-2011-0dG, Swedish Cancer Society, Swedish Research Council, MAS: We would like to thank Famida Zulkiﬁ and Ms. Moey for assistance in patient recruitment, data collection, and sample preparation. The Malaysian Ovarian Cancer (Giles, R.M. Glasspool by research grants from the Malaysian Ministry of Higher Education (UM-C/HIR/MOHE/06) and charitable funding from Cancer Research Initiatives Foundation; MCCS: MCCS cohort recruitment was
funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553, and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database; MOE: The Total Cancer Control Protocol and the Collaborative Data Services and Tissue Core Facilities at the H. Lee Moffitt Cancer Center & Research Institute, an NCI designated Comprehensive Cancer Center (P10-C0706292), Merck Pharmaceuticals and the state of Florida; NHS: The NHS/NIIH 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; OPL: Members of the OPAL Study Group [http://opalsudy.qimrberghofer.edu.au/]; RPC: NIH [R01 CA159981, R01 CA162684]; SEK: SEARCH team, Craig Luccarini, Caroline Baynes, Don Conroy; SIS: The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-E0044005 and Z01-E0049033); SON: National Health Research and Development Program, Health Canada, grant 6613-SEA; SIS: Baynes, Don Conroy; Study Group (http://opalstudy.qimrberghofer.edu.au/); RPC: NIH [R01 CA0900742], DRE: Ovarian Cancer Research Fund; DOW: NIH R01-CA112523 and R01-CA87538; EMC: Dutch Cancer Society (EMC 2014-6699); EPC: The coordination of EPC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Lung Cancer Study Group, Institute Gustave Roussy, Montepaone, Centre de l’Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM, France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF, Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-ABC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKI); IRK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); ERC-2009-AdG 232997 and NordForsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); Health Research Fund (HS), P11/00061 to Granada, P11/01162 to EPIC-Murcia, Regional Governments of Andalucia, Asturias, Basque Country, Murcia and Navarra, ICSICH RETIC (BD06/0020); Spain; Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västbotnen (Sweden); Cancer Research UK (14136 to EPIC-Norfolk), the Leu V S. Smith Foundation; Foundation lVTS-39420; UHIN: Princess Margaret Cancer Centre Foundation-Fund d’IGF 2006 for the Care; UKO: We particularly thank I. Jacobs, M.Widschwendter, E. Wozniak, A. Ryan, J. Ford, and N. Balogun for their contribution to the study; in addition, the following data were supplied by the study: the Whole Genome Atlas Pilot Project, established by the National Cancer Institute (http://www.ncbi.nlm.nih.gov/); the Ovarian Cancer Research Fund thanks to donations by the family and preparation of the manuscript.

This project was partially supported by the development fund from the Department of Medicine at Vanderbilt University Medical Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations from the family and friends of Kathryn Sladek Smith (PDP/RPCI 07). The scientific development and funding for this project were, in part, supported by the US National Cancer Institute CA83814 Post-DOC/AS Initiative (I11-CA148111). This study is part of a network of data generated by the Cancer Genome Atlas project (CCAG), a national resource of the National Cancer Institute, supported by the Office of the Director, National Institutes of Health, and the NIH Roadmap for Medical Research. This research has been cofounded by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program ‘Education and Life-long Learning’ of the National Strategic Reference Framework (NSRF)- Research Funding Program of the General Secretariat for Research and Technology: #10_19_1981. The Authors declare no competing interests.

This project was partially supported by the development fund from the Department of Medicine at Vanderbilt University Medical Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations from the family and friends of Kathryn Sladek Smith (PDP/RPCI 07). The scientific development and funding for this project were, in part, supported by the US National Cancer Institute CA83814 Post-DOC/AS Initiative (I11-CA148111). This study is part of a network of data generated by the Cancer Genome Atlas project (CCAG), a national resource of the National Cancer Institute, supported by the Office of the Director, National Institutes of Health, and the NIH Roadmap for Medical Research. This research has been cofounded by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program ‘Education and Life-long Learning’ of the National Strategic Reference Framework (NSRF)- Research Funding Program of the General Secretariat for Research and Technology: #10_19_1981. The Authors declare no competing interests.
Cancer Institute of New Jersey; NOR: Helse Vest, The Norwegian Cancer Society; The Research Council of Norway; NTH: Radboud University Medical Center; OPL: National Health and Medical Research Council (NHMRC) of Australia (APP1025142) and Brisbane Women's Club; ORE: OHSU Foundation; OVA: Canadian Institutes of Health Research grant (MOP-66527) and by NIH/NCI 1R01CA166969-01A1; PLC: Intramural Research Program of the National Cancer Institute; POC: Pomeranian Medical University; POL: Intramural Research Program of the National Cancer Institute; PVD: Canadian Cancer Society and Cancer Research Society GifPePEC Program; RBH: National Health and Medical Research Council of Australia; RMH: Cancer Research UK; Royal Marsden Hospital; RPC: NIH (PS0 CA159981, R01CA126841); SEA: Cancer Research UK (C490/A10119 C490/A10124); UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge; SIS: NIH, National Institute of Environmental Health Sciences, 201 ES044005 and 201 ES049033; SMC: The Swedish Research Council; SNNC: National Health Research and Development Program, Health Canada, grant 6613-1415-53; SRO: Cancer Research UK (C356/13A086, C356/ A6689) and Imperial Experimental Cancer Research Centre (C1312/A15859); STA: NIH grants L01 CA71966 and L01 CA69417; SWE: Swedish Cancer Foundation, WeCanCureCancer and arkKampMostCancer foundation; NIH (NCI) grant R01-CA070867; TBO: NIH (R01 CA061442-A2), American Cancer Society (CRTG-00-196-01-CCE), Department of Defense (DAMD17-98-1-8659), Celma Mastery Ovarian Cancer Foundation; TOR: NIH grants R01 CA063678 and R01 CA063682; UCI: NIH R01-CA058860 and the Leon V Smith Foundation grant CVS-39420; UHN: Princess Margaret Cancer Centre Foundation; UK: The UKOPS study was funded by The Eve Appeal [The Oak Foundation] and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre; UKR: Cancer Research UK [C490/A66187], UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge; UBC: P01CA17054, P30CA14089, R01CA61132, N01PC67010, R03CA113148, R03CA115195, N01CO25403, and California Cancer Research Program (00-01389V-2017, 2H2020); VAN: BC Cancer Foundation, VGH & UBC Hospital Foundation; VTI: NIH R05-CA154337; WMH: National Health and Medical Research Council of Australia, Enabling Grants 310670 and 628/093.

References

8. Jansen RC, Norling G, Ludemann C, Drobyski WR, van der Ro初次, Ed Dicks, Craig Luccarini and
9. The Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland.

We are grateful to the family and friends of Kathryn Sladek Smith for their generous support of the Ovarian Cancer Association Consortium through their donations to the Ovarian Cancer Research Fund. The OncoArray and COGS genotyping projects would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Marjorie J. Riggs (OCAC). Rosalind A. Feles, Douglas F. Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Antonis Antoniou, Lesley McGuftog, Fergus Couch and Ken O'efft (CIMBA), Alison M. Dunning, Andrew Lee, and Ed Dirks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessler, Francois Bacoit, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Genome Quebec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge Nordestgaard, and the staff of the Copenhagen DNA Laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. We pay special tribute to the contribution of Professor Brian Henderson to the CAN-080 consortium and to Olga M. Similinkova for her contribution to CIMBA and for her part in the initiation and coordination of GEMO until she sadly passed away on the 30th June 2014. We thank the study participants, doctors, nurses, clinical and scientific collaborators, health care providers and health information sources who have contributed to the many studies contributing to this manuscript.

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

Received August 30, 2018; revised October 16, 2018; accepted December 6, 2018, published first December 17, 2018.


Genetic Data from Nearly 63,000 Women of European Descent Predicts DNA Methylation Biomarkers and Epithelial Ovarian Cancer Risk

Yaohua Yang, Lang Wu, Xiang Shu, et al.

Cancer Res 2019;79:505-517. Published OnlineFirst February 1, 2019.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-18-2726

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2018/12/15/0008-5472.CAN-18-2726.DC1

Cited articles
This article cites 52 articles, 7 of which you can access for free at:
http://cancerres.aacrjournals.org/content/79/3/505.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/79/3/505. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.