Published in final edited form as:
Genet Epidemiol. 2014 July ; 38(5): 457–466. doi:10.1002/gepi.21815.

Genome-wide investigation of regional blood-based DNA methylation adjusted for complete blood counts implicates BNC2 in ovarian cancer

Stacey J. Winham[1,§], **Sebastian M. Armasu**[1], **Mine S. Cicek**[1], **Melissa C. Larson**[1], **Julie M. Cunningham**[2], **Kimberly R. Kalli**[3], **Brooke L. Fridley**[4,§], and **Ellen L. Goode**[1]

[1]Department of Health Sciences Research, Mayo Clinic, Rochester MN

[2]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester MN

[3]Department of Oncology, Mayo Clinic, Rochester MN

[4]Department of Biostatistics, Kansas University Medical Center, Kansas City KS

Abstract

Due to its potential as a biomarker for early cancer detection, blood-based DNA methylation (DNAm) is of interest in cancer research. Specifically, highly predictive mechanisms for early detection of epithelial ovarian cancer (EOC) are desired, so previous studies have compared DNAm between EOC cases and controls. However, case-control studies are confounded by the distribution of white blood cell types through an immune response induced by the cancer. Rather than determining the distribution of the cell types manually or investigating isolated cell types, an alternative approach involves the use of complete blood count (CBC), which is routinelycollected. In the analysis of an EOC case-control study of DNAm, we incorporate CBC measures to adjust for this confounding and compare DNAm between 242 EOC cases and 181 age-matched controls (assayed on the Illumina Infinium HumanMethylation27 or HumanMethylation450 Beadchips), at both the individual CpG and CpG island levels. We found that adjustment for leukocyte distribution using CBC measurements dramatically reduced confounding, with 62 single CpG sites found to be associated with EOC status after adjustment ($p<5E-8$). Additionally, regional DNAm was assessed by applying principal components analysis to CpG islands. The top associated CpG island ($p=7E-6$) was located in the promoter/transcription start site of the human basonuclin 2 gene (*BNC2*), a known susceptibility gene for EOC risk identified through GWAS. Follow-up studies are necessary to establish the role of *BNC2* in blood-based DNA and EOC, including prospective studies to validate this region as a potential biomarker and predictor of EOC susceptibility.

Keywords

Complete blood count; etiology; epigenetics; CpG island; principal components

[§]Corresponding Authors: Dr. Stacey Winham, 200 First Street SW, Rochester, MN 55905, (507) 266-3662, winham.stacey@mayo.edu, Dr. Brooke L. Fridley, 3901 Rainbow Blvd, Kansas City, KS 66160, (913) 945-5039, bfridley@kumc.edu.

Introduction

Familial ovarian cancer is under strong genetic control, with many known high-penetrance genetic risk factors, such as the rare BRCA1/2 mutations [Szabo and King 1997]. Similarly, a number of common, low-penetrance susceptibility variants have been identified for invasive epithelial ovarian cancer (EOC) [Bolton, et al. 2010; Goode, et al. 2010; Pharoah, et al. 2013]; however, these known variants explain only a fraction of the genetic variation in EOC risk, requiring the consideration of the role of epigenetics and other regulatory elements in the development of EOC. Recently, many studies of epigenetics, such as DNAm, have been completed that add to the understanding of cancer development, risk and progression [Feinberg, et al. 2006; Irizarry, et al. 2009], including ovarian cancer [Shen, et al. 2013]. Many such studies have focused on methylation patterns in ovarian tumor tissue, demonstrating a clear role of DNAm in EOC tumors and facilitating improved understanding of cancer biology [Bell, et al. 2011; Cicek, et al. 2013].

While the role of blood-based DNAm is less understood, it nonetheless has become a subject of interest in cancer research, given its potential as a biomarker for noninvasive early detection [Widschwendter, et al. 2008]. Currently, there are few highly predictive mechanisms for early detection of EOC, with the exception being sequential screening with carbohydrate antigen 125 via the Risk of Ovarian Cancer Algorithm [Lu, et al. 2013; Skates, et al. 2001]. Approximately 75% of women are not diagnosed until advanced stage III or IV disease, when prognosis is poor with low rates of long-term survival (<30%) [Hennessy, et al. 2009]. Therefore investigation of the impact of blood-based DNAm is of interest, due to its potential as an easy-to-use biomarker. For instance, a previous study of blood-based DNAm at approximately 27,000 CpG sites in 261 subjects identified a DNA methylation signature based on 100 CpG sites that was associated with EOC case-control status [Teschendorff, et al. 2009]. This suggests that blood-based DNAm could be a powerful biomarker for EOC risk prediction and early disease detection, although this study did not adjust for an important confounding factor of blood cell type.

Specifically, EOC risk is influenced by a number of epidemiological factors in addition to the genetic component, including reproductive history such as parity and oral contraceptive use [Adami, et al. 1994; Whittemore, et al. 1992], lifestyle factors such as obesity and smoking [Jordan, et al. 2006; Leitzmann, et al. 2009], and demographic factors [Hunn and Rodriguez 2012]. Unlike germline sequence variation, the epigenome is undergoing constant modification as a result of the influence of many of these same external, environmental factors; for instance, aging, smoking, and alcohol consumption are all associated with altered DNAm in gene promoter regions [Breitling, et al. 2011; Christensen, et al. 2009; Philibert, et al. 2012]. This provides a natural model for the role of epigenetics in EOC disease etiology, although such a fact also elicits concerns about the high potential for confounding variables; therefore known confounding exposures such as smoking and alcohol consumption are routinely included in analyses of DNAm. It is known that markers of systemic inflammatory response are prognostic indicators for cancer outcomes[Fox, et al. 2013] and distributions of leukocyte cell types differ between cancer cases and controls [Yamanaka, et al. 2007], but recently it has been shown that the particular mixture of leukocyte cell types from which the DNA is collected is also associated with DNAm at

specific CpG sites [Adalsteinsson, et al. 2012; Houseman, et al. 2012; Koestler, et al. 2012]. Because the onset of cancer may cause an immune response that alters the distribution of white cell types[Hanahan and Weinberg 2011; Mantovani, et al. 2008], it is important to account for this cell type distribution in analyses, which was not done in the previous study of epigenome-wide DNAm and EOC status [Teschendorff, et al. 2009]. Ideally, analysis should be conducted at the level of the cell type, but this requires that cell types be isolated through techniques such as magnetic activated cell-sorting prior to DNA extraction. However, peripheral blood mononuclear cells are preferred as a source of DNA because of the ease of collection and use; such cell type isolation complicates this use, particularly in a clinical setting. An alternative clinical measure to assess the leukocyte distribution is the complete blood count (CBC), routinely collected at time of blood draw. Recently, CBC measures were used as covariates in a study of DNAm of CpG sites within candidate CpG islands to account for cellular heterogeneity [Adalsteinsson, et al. 2012], although such an approach has yet to be applied at the level of the epigenome.

Furthermore, like DNA sequence variation, DNAm is correlated on the regional level, where nearby CpG sites tend to have similar methylation levels. Such correlation structure could be leveraged to improve power if correlated CpGs are viewed as a set. Typically DNAm has been examined either at the individual CpG sites or via a full methylation profile, rather than an intermediate focus on DNAm at the regional level (for example, CpG islands); the exception is the application of the statistical technique of ‘bump-hunting’ to examine regional methylation effects for quantitative traits [Jaffe, et al. 2012]. In contrast, many set-based approaches have been developed and successfully applied to disease studies of DNA sequence variation at the gene-level [Chapman and Whittaker 2008; Gauderman, et al. 2007; Lehne, et al. 2011], and such approaches could also prove valuable in studies of DNAm in order to reduce multiple testing burdens and utilize prior biological knowledge.

The goal of this study is to characterize associations with DNAm and EOC status at CpG island regions, while accounting for the confounding effects of cell type distribution. This study represents the first to incorporate CBC measures into the investigation of the relationship between blood-based DNAm and ovarian cancer risk across the epigenome, with a novel focus on associations with DNAm at the regional level of CpG islands.

Materials and Methods

Sample Characteristics and CBC Measurements

Ovarian cancer cases were women age 20 years or older recruited within one year of diagnosis of pathologically confirmed primary epithelial ovarian, fallopian tube, or primary peritoneal cancer at Mayo Clinic between 2000 and 2009. Cases provided a peripheral blood sample prior to the start of chemotherapy. Controls were matched to cases on age (within one year) and area of residence, and were recruited from women seen at Mayo Clinic for general medical examinations. All subjects were of European descent, residing in a six-state region of the Upper Midwest. Table 1 summarizes the clinical and demographic characteristics of 242 cases and 181 controls that were subsequently analyzed. This study was approved by the Mayo Clinic Institutional Review Board and all participants provided written informed consent.

The subset of subjects in this study was previously assayed as part of a study of DNA methylation on 336 EOC and 398 control subjects using either the Illumina Infinium HumanMethylation27 or HumanMethylation450 Beadchips [Fridley, et al. Under Review]. From this larger pool of subjects, pre-treatment CBC measurements were extracted from the medical record when available, resulting in 242 cases and 181 controls with available CBC measurements—including total leukocytes, consisting of lymphocytes, neutrophils, monocytes, basophils, and eosinophils (Figure 1). In general, subjects with available CBC measurements were not clinically different from those without, except that such measures were available for a larger proportion of EOC cases (Supplemental Table 1). Only individuals with CBC measurements obtained within 2 weeks of DNA blood draw were eligible for analysis, in order to ensure cell measurements reflect similar distributions to the DNA. Because lymphocytes, neutrophils, monocytes, basophils, and eosinophils are subgroups of leukocytes (Figure 1), subjects with total leukocyte counts not equal to the sum of the other cell types were ineligible. Total leukocyte counts consisted primarily of neutrophils followed by lymphocytes (Table 1), which were inversely correlated (Supplemental Table 2). Distributions of total leukocytes differed for EOC cases and controls, with cases having higher neutrophils and lower lymphocytes (Supplemental Figure 1).

Methylation Arrays

All subjects were previously assayed on either the Illumina Infinium HumanMethylation27 (N=523) or HumanMethylation450 Beadchips (N=313) (i.e. 27K or 450K array). Subjects assayed on the HumanMethylation27 were assayed in one of two batches (N=175 and 348). For each assay, DNA methylation values were scored as beta values ranging from 0 to 1 (unmethylated to methylated). Beta values were obtained for 27,578 CpG probes from the 27K array and 485,577 probes from the 450K array. CpG probes present on the 27K array provide low density epigenome coverage, with only a single probe per genomic region, whereas the 450K array offers denser coverage with multiple CpG probes per region, with a focus on CpG islands (genomic regions with a high frequency of CpG sites usually near gene promoters), shores (regions flanking a CpG island), and shelves (regions flanking a CpG shore). Many of the probes present on the 27K array are also present on the 450K array (96%), and this overlap allows the opportunity for independent replication for this set.

Quality Control Procedures

Quality control of the sample data, including CBC measurements, results in a sample size of 153 EOC cases and 107 controls assayed on the 27K array and 89 EOC cases and 74 controls assayed on the 450K array (total N=423). Quality control was carried out separately for each batch. CEPH controls, positive and negative bisulfite modified controls, and sample duplicates were included in each batch, and 20 duplicates were also included across batch to assess data quality. Samples that were determined to be outliers based on bisulfite conversion ratios and call rates were excluded (detection p-value of 0.05).

Probes detected in less than 70% of samples and with beta values beyond four standard deviations from the mean in bisulfite negative controls were excluded. Probes residing on the Y chromosome were also removed, as were probes known to map to non-specific

genomic locations on the 27K and 450K arrays [Chen, et al. 2011; Chen, et al. 2013]. After quality control, 23,145 CpG probes were available for analysis on the 27K array and 413,425 CpG probes on the 450K array. This includes 22,278 probes present on both arrays that can be compared across arrays for validation.

Through principle components analysis, we observed a plate effect in the 450K array batch and a chip within plate effect in the 27K array batches. We performed normalization to adjust for plate effects (450K array) and chips within plate (27K array). For each CpG site, the residuals from a linear model of the logit-transformed beta values were computed. The logit-transformed locus mean was added back to the residuals, and then transformed to the 0 to 1 scale to obtain an adjusted beta value for each CpG site. For further across study/ experiment normalization, we applied an empirical Bayes method via 'Combat' in order to remove batch effects and combine the two batches assayed with the 27K array [Johnson, et al. 2007].

Statistical Analysis

When examining associations with EOC disease status and blood-based DNA methylation, it is critical to properly adjust for white blood cell distribution to eliminate confounding. We aim to approximate this distribution using CBC measures (total leukocytes, as well as neutrophils, lymphocytes, monocytes, basophils, and eosinophils). We examined the relationship between each of the cell types and case-control status, both separately and jointly (with logistic regression). To account for individual differences in count, we utilized the proportion of each cell type out of total number of leukocytes. Due to the high correlation among CBC measures (Supplemental Table 2), it is unnecessary to include all proportions for adjustment. The proportion of neutrophils (and inversely the proportion of lymphocytes) was most strongly associated with case-control status, and hence proportion of neutrophils was chosen as a summary measure of CBC distribution [Adalsteinsson, et al. 2012]. Other important adjustment factors included age at first birth, smoking status, alcohol use, state of residence, and time of study enrollment (Table 1) [Fridley, et al. Under Review].

Next we examined associations between DNAm and EOC status. Due to departures from normality in the distribution of adjusted beta values, we applied an arc sin root transformation to each CpG site to reduce skewness, (i.e. $\arcsin(\sqrt{m})$, where m is the adjusted beta value). This transformation has been suggested previously [Adalsteinsson, et al. 2012; Koestler, et al. 2012], and for the top-ranking locus, we perform sensitivity analyses to assess the impact of this transformation. We first performed analysis at the individual CpG level, separately for subjects assayed on the 27K and 450K arrays, including the CpG probes unique to each platform. For each CpG, we fit a linear model to the transformed methylation beta value as a function of case-control status, adjusted for age at first birth, smoking status, alcohol use, state of residence, and time of study enrollment. To assess the confounding effect of inflammation, we analyzed the data with and without adjustment for proportion of neutrophils. We conducted single CpG analyses separately in both platforms to allow for independent replication across arrays, and subsequently performed a random-effects meta-analysis via the DerSimonian-Laird method in order to

combine coefficient estimates across both sets (for the common probes). For each probe, heterogeneity across arrays was also assessed using Cochrane's Q statistic. Reported p-values are not adjusted for multiple testing, but are considered significant if p<5-8 in the meta-analysis.

We also investigated case-control associations with DNAm at the level of the CpG island. Because CpG probes are sparse across islands on the 27K array, analysis was restricted to those subjects assayed on the denser 450K array (89 cases, 74 controls). We grouped CpG sites into regions based on location in a CpG island, shore (+/- 2KB), or shelf (+/- 4KB), based on Genome Build 37. For a given CpG island region, the island, shore, and shelf were analyzed together. For all CpG sites within a defined island region (including the shore and shelf), we obtained a summary measure of the regional methylation using principle components analysis (PCA). PCA is often utilized for set-based analysis of SNPs aggregated at the gene-level, and has been shown to have high power in this setting [Gauderman, et al. 2007], and the extension to DNAm is straight-forward. For each region, we modeled case-control status as a function of the first principal component (explaining the largest proportion of variation) with logistic regression, adjusted for proportion of neutrophils, age at first birth, smoking status, alcohol use, state of residence, and time of study enrollment. Due to the large number of components that would be necessary to explain 80% variation relative to the sample size, we restricted our analysis to the first principal component only to ensure stable estimates, and report the percent variation explained. Analysis was conducted using R statistical software (version 2.14.0), and meta-analyses of single CpG estimates were performed using the R package 'rmeta' (http://cran.us.r-project.org/) and PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/).

Results

Before adjustment for the proportion of neutrophils, a large number of individual CpGs appear to be strongly associated with disease status, in both independent sets (27K and 450K) and in the meta-analysis, with many probes having p-values in the range 1E-10<p<1E-50 (Figure 2), clearly demonstrating the impact of confounding. After adjustment for proportion of neutrophils, the inflation of p-values is dramatically reduced (Figure 3). Results for the top CpGs in the meta-analysis (representing CpGs that replicate across arrays with p<5E-8) are presented in Table 2, along with the individual results for the 27K and 450K subsets; for probes exhibiting evidence of potential associations, results from the 27K and 450K subsets are positively correlated (Supplemental Figure 2). Sixty-two probes displayed evidence for association with EOC disease status with replication at level p<5E-8 in the meta-analysis; 29 of these probes reside in CpG islands or shores. The top associated CpG site (probe cg11738543) was on chromosome 12, which is located in a CpG island within an intron of suppressor of cytokine signaling 2 (*SOCS2*), and shows evidence of hypo-methylation in cases (regression coefficient from meta-analysis =-0.045, meta-analysis p-value =9E-14).

In general, most CpG probes were hypo-methylated in EOC cases in both the 27K and 450K array subsets. CpG probes unique to the 27k array (867 probes) and the 450K array (391,147

probes) were not available for possible replication and meta-analysis, but results are presented for probes with p<5E-8 (Supplemental Tables 3 and 4).

CpG sites assayed on the 450K array were grouped into 25,607 CpG island regions consisting of the shore and shelf regions surrounding the CpG island. From the region-based test utilizing PCA, 6 CpG island regions demonstrated evidence of association with EOC disease status at the level p<1E-5 (Table 3). The top region is a CpG island on chromosome 9, 'chr9:16870123-16872020', which included 9 CpG probes (p=7E-6). Although this region does not meet a strict Bonferroni criterion for statistical significance (p<2E-6), this region is located in the promoter/transcription start site of the human basonuclin 2 gene (*BNC2*), which encodes the zinc finger basonuclin 2 protein (Figure 4). The top ranking CpGs within this region are cg04389426 and cg05674150 (p<5E-6, Supplemental Table 5). Due to the sparsity of CpGs within the *BNC2* island region assayed on the 27K array, replication of this island-level analysis was not able to be completed. Sensitivity analysis was conducted for the CpGs within this region to assess the impact of the arc sin root transformation on the DNAm beta values, and the results were unchanged (Supplemental Table 6).

Discussion

In this first study to incorporate information on complete blood counts into the epigenome-wide analysis of EOC status in blood-based DNA, we examined associations with DNAm at each CpG probe in two cohorts (assayed on the Illumina Infinium HumanMethylation27 or HumanMethylation450 Beadchips), combining evidence of EOC association across both sample sets using random-effects meta-analysis to identify replicating CpG sites. These results emphasize the need to adjust for the potential confounding effect of leukocyte cell type distribution when examining blood-based DNA methylation, as distributions in leukocyte cell type may reflect a possible immunological response or consequence of EOC. In particular, in this study, a large number of CpG sites displayed strong associations with EOC status that disappeared after adjustment based on the proportion of neutrophils. Although such confounding effects have been documented in previous studies [Houseman, et al. 2012; Koestler, et al. 2012], this work establishes the utility of complete blood counts as an alternative covariate measure to isolated cell type, and provides regions of DNAm associated with EOC disease status.

Furthermore, the top results based on our meta-analysis typically differ from those previously reported for EOC risk prior to adjustment for cell type distribution [Teschendorff, et al. 2009], although some top CpG sites are also highly ranked in our analysis. When comparing the 62 probes observed here at p<5E-8 to the top 115 CpG probes with p<5E-8 in Teschendorff et al., we observe an overlap of 21 CpG sites, including cg03330678 within septin 9 (*SEPT9*), which demonstrates hypo-methylation in cases in both our data (p=8E-13) and the previous data (p =3E-9). *SEPT9* is a member of the septin family involved in cytokinesis, and notably, DNAm in *SEPT9* is associated with colorectal cancer [deVos, et al. 2009], and *SEPT9* has previously been implicated in ovarian tumorigenesis [Scott, et al. 2006]. However, the methylation signature identified by Teschendorff et al. also implicates genes differentially expressed across blood cell types, possibly as a result of confounding with cell-type effects. Rather, our meta-analysis yielded 62 CpG sites at the level of p<5E-8

after adjustment for cell type distribution, and the top CpG result is located in a CpG island within an intron of suppressor of cytokine signaling 2 (*SOCS2*). *SOCS2*, a member of the *SOCS* gene family, is a negative regulator of cytokine receptor signaling, acting on the Janus kinase/signal transducer and activation of transcription pathway (JAK/STAT pathway) [Rico-Bautista, et al. 2006]. Members of the *SOCS* gene family have been implicated in various cancers, and in particular *SOCS2* has been recently linked to both prostate and breast cancer [Farabegoli, et al. 2005; Haffner, et al. 2007; Hendriksen, et al. 2006; Iglesias-Gato, et al. 2013; Leung, et al. 2003; Zhu, et al. 2013].

In addition to consideration of associations with DNAm at the level of an individual CpG site, we also conducted a novel regional-level exploration of DNAm and EOC status by grouping CpG probes into CpG islands. By condensing CpGs from individual sites into broader regions, we are able to gain power by leveraging correlation across CpG sites and reducing the total number of analyses performed [Gauderman, et al. 2007; Jaffe, et al. 2012], as well as explore an additional epigenetic model where DNAm is assumed to act on the level of the CpG island. In general, the island-level regional analysis represents a complementary approach to the single CpG analysis, with different regions and sites implicated in each: the islands containing highly ranked CpGs were not highly ranked at the regional level, and vice versa. This emphasizes the importance of considering both types of approaches to target different underlying biology. For instance, single CpG sites could be missed if only considered at the regional level, and conversely, important regions could be over-looked by relying on single high-ranking CpGs rather than the strength of combined evidence. This is illustrated specifically in our study, where the island-level analysis implicates the promoter region of *BNC2*; while this region was not included on the 27K array and therefore not included in the meta-analysis, no individual CpGs within this region were ranked among the top of the CpG sites present on the 450K array (Supplemental Table 3).

The promoter region of *BNC2* is of particular interest in regards to EOC. The human basonuclin 2 gene encodes the zinc finger basonuclin 2 protein, a possible transcription factor, which is highly expressed in reproductive tissue and may be involved in the differentiation of spermatozoa and oocytes [Romano, et al. 2004]. Notably, a DNA sequence variant in locus 9p22 (within *BNC2*) has been confirmed as a susceptibility locus for EOC [Goode, et al. 2010; Song, et al. 2009]. In particular, this SNP is located between the coding regions of *BNC2* and *CNTLN*, 42 KB upstream of the CpG island within the *BNC2* promoter identified here. Furthermore, the SNPs in the 9p22 locus have been associated with abnormal ovarian ultrasounds [Wentzensen, et al. 2011], and lower *BNC2* expression has been demonstrated in EOC cell cultures compared to normal ovarian cell lines [Goode, et al. 2010]. While *BNC2* sequence variation and gene expression play a role in EOC risk, association with DNAm was previously unknown. Our study implicates a CpG island within the promoter region of *BNC2* for association with EOC status; although this CpG island did not quite reach the criterion for statistical significance (p=7E-6 vs. 2E-6), prior evidence for this region are convincing. Although the role of DNAm within *BNC2* in EOC risk is unknown, we hypothesize that it may mediate the effect on gene expression, based on the observed differences in expression between EOC and normal ovarian tissue [Goode, et al.

2010]. Further studies should investigate this hypothesis by examining associations between DNAm and gene expression, particularly within ovarian tissue. Furthermore, in this study, we detected the *BNC2* region through an association with blood-based methylation (rather than in the tumor), and therefore future studies are necessary to entangle the role of *BNC2* in EOC by investigating both germline and tumor DNA simultaneously.

In general, EOC cases were hypo-methylated at most CpG sites across the epigenome as compared to controls, consistent with observations from other studies of global DNAm and cancer risk [Brennan and Flanagan 2012]. However, within CpG island promoter regions, methylation is known to hinder transcription initiation (and thus block expression), and therefore cancer cases are typically hyper-methylated in the promoter region of tumor suppressor genes [Jones and Baylin 2002]. But such observations have primarily been investigated in tumor tissue rather than via blood-based DNA. In this study of blood-based DNA, even CpGs within islands displayed hypo-methylation of EOC cases; a similar pattern of hypo-methylation in CpG island regions has also been observed in other EOC studies focusing on blood-based DNA [Teschendorff, et al. 2009].

Although findings from this study are intriguing, results are limited primarily by sample size. Because of a reduced sample size resulting from inclusion of CBC adjustment (as CBC measurements within two weeks of blood draw were not available for all subjects) in addition to analyzing the 27K and 450K subsets separately, power to detect differentially methylated regions was limited. However, our hope is that this loss in power due to low sample size is offset by the reduced potential for false positive findings due to the critical covariate adjustment, regional focus, and the inclusion of both cohorts for replication via meta-analysis. Furthermore, previous studies of blood-based methylation and EOC disease status have had similar sample sizes [Koestler, et al. 2012; Teschendorff, et al. 2009].

While our analysis of DNAm on the regional level is novel, it is limited by how the region is defined. In this study, we chose regions based on the biological definition of the CpG island in order to incorporate potential biological relationships (i.e. regulation of gene expression via promoter methylation) to improve the probability of eliciting associations of functional relevance. However, more agnostic approaches are also possible, such as a sliding window approach, or a data-driven definition of the region [Jaffe, et al. 2012]. Furthermore, we included CpG shores and shelves within the scope of the CpG island, although these regions could also be examined separately. However, in this study, analysis of CpG shores and shelves separately from the islands does not impact the results—the *BNC2* promoter island remains the top ranked region (data not shown).

Results from this study are also limited in the scope of interpretation. As data were obtained from a retrospective case-control study rather than a prospective cohort design, associated regions cannot be interpreted as indicators of EOC susceptibility/risk. In addition, as blood was drawn upon diagnosis for EOC cases, any case-control differences may reflect changes in methylation due to the cancer itself, and do not necessarily reflect biomarkers that would be predictive for risk prediction or early detection. In fact, it is possible that circulating tumor cells present in the blood could be driving the observed associations; for example, lower *BNC2* gene expression has previously been observed in EOC tumor cells[Goode, et al.

2010]. However, the percentage of tumor contamination is thought to be extremely low (<60 cells/7.5mL), and therefore effect sizes of tumor methylation would need to be extremely large for a contaminated effect to be detectable within blood-based DNA [Allard, et al. 2004]. Nonetheless, follow-up studies in prospective cohorts should be conducted to determine whether associated regions could represent blood-based markers of disease risk, or rather are simply markers of existing disease.

In summary, this study represents the first to incorporate data on complete blood counts into the investigation of EOC status in blood-based DNA methylation across the epigenome, reducing false positive associations by providing evidence of replication through the use of meta-analysis and controlling for the confounding effects of the inflammatory response to EOC. Furthermore, this is also the first study to examine the role of methylation in EOC at the regional level, identifying a CpG island promoter region in *BNC2*, a known EOC susceptibility gene. Future studies should further explore the role of DNA methylation in *BNC2*, including its relation with sequence variation and gene expression, as well as its potential role as a blood-based biomarker of EOC risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this study was provided by the Fred C. and Katherine B. Andersen Foundation and the National Institute of Health (P30 CA168524, P50 136393, R01 CA122443, P20 GM103418, R21 GM86689, K12 HD65987), as well as Genetic Associations and Mechanisms in Oncology (GAME-ON), a NCI Cancer Post-GWAS Initiative (U19-CA148112).

References

Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV, Gudnason V. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS One. 2012; 7(10):e46705. [PubMed: 23071618]

Adami HO, Hsieh CC, Lambe M, Trichopoulos D, Leon D, Persson I, Ekbom A, Janson PO. PARITY, AGE AT FIRST CHILDBIRTH, AND RISK OF OVARIAN-CANCER. Lancet. 1994; 344(8932): 1250–1254. [PubMed: 7967985]

Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, Tibbe AGJ, Uhr JW, Terstappen L. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clinical Cancer Research. 2004; 10(20):6897–6904. [PubMed: 15501967]

Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, Dhir R, DiSaia P, Gabra H, Glenn P, et al. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474(7353):609–615. [PubMed: 21720365]

Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, Sher T, Gentry-Maharaj A, Wozniak E, Tsai YY, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nature Genetics. 2010; 42(10):880–+. [PubMed: 20852633]

Breitling LP, Yang RX, Korn B, Burwinkel B, Brenner H. Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication. American Journal of Human Genetics. 2011; 88(4):450–457. [PubMed: 21457905]

Brennan K, Flanagan JM. Is There a Link Between Genome-Wide Hypomethylation in Blood and Cancer Risk? Cancer Prevention Research. 2012; 5(12):1345–1357. [PubMed: 23135621]

Chapman J, Whittaker J. Analysis of multiple SNPs in a candidate gene or region. Genetic Epidemiology. 2008; 32(6):560–566. [PubMed: 18428428]

Chen YA, Choufani S, Ferreira JC, Grafodatskaya D, Butcher DT, Weksberg R. Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. Genomics. 2011; 97(4):214–22. [PubMed: 21211562]

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013; 8(2):203–9. [PubMed: 23314698]

Christensen BC, Houseman EA, Marsit CJ, Zheng SC, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, et al. Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. Plos Genetics. 2009; 5(8)

Cicek MS, Koestler DC, Fridley BL, Kalli KR, Armasu SM, Larson MC, Wang C, Winham SJ, Vierkant RA, Rider DN, et al. Epigenome-wide ovarian cancer analysis identifies a methylation profile differentiating clear-cell histology with epigenetic silencing of the HERG K+ channel. Hum Mol Genet. 2013; 22(15):3038–47. [PubMed: 23571109]

deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, Steiger KV, Grutzmann R, Pilarsky C, Habermann JK, et al. Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. Clin Chem. 2009; 55(7):1337–46. [PubMed: 19406918]

Farabegoli F, Ceccarelli C, Santini D, Taffurelli M. Suppressor of cytokine signalling 2 (SOCS-2) expression in breast carcinoma. J Clin Pathol. 2005; 58(10):1046–50. [PubMed: 16189149]

Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nature Reviews Genetics. 2006; 7(1):21–33.

Fox P, Hudson M, Brown C, Lord S, Gebski V, De Souza P, Lee CK. Markers of systemic inflammation predict survival in patients with advanced renal cell cancer. Br J Cancer. 2013; 109(1):147–53. [PubMed: 23778526]

Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. Genet Epidemiol. 2007; 31(5):383–95. [PubMed: 17410554]

Goode LL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, Widschwendter M, Vierkant RA, Larson MC, Kjaer SK, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nature Genetics. 2010; 42(10):874–+. [PubMed: 20852632]

Haffner MC, Petridou B, Peyrat JP, Revillion F, Muller-Holzner E, Daxenbichler G, Marth C, Doppler W. Favorable prognostic value of SOCS2 and IGF-I in breast cancer. BMC Cancer. 2007; 7:136. [PubMed: 17651480]

Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011; 144(5):646–674. [PubMed: 21376230]

Hendriksen PJ, Dits NF, Kokame K, Veldhoven A, van Weerden WM, Bangma CH, Trapman J, Jenster G. Evolution of the androgen receptor pathway during progression of prostate cancer. Cancer Res. 2006; 66(10):5012–20. [PubMed: 16707422]

Hennessy BT, Coleman RL, Markman M. Ovarian cancer. Lancet. 2009; 374(9698):1371–82. [PubMed: 19793610]

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012; 13:86. [PubMed: 22568884]

Hunn J, Rodriguez GC. Ovarian Cancer: Etiology, Risk Factors, and Epidemiology. Clinical Obstetrics and Gynecology. 2012; 55(1):3–23. [PubMed: 22343225]

Iglesias-Gato D, Chuan YC, Wikstrom P, Augsten S, Jiang N, Niu Y, Seipel A, Danneman D, Vermeij M, Fernandez-Perez L, et al. SOCS2 mediates the cross talk between androgen and growth hormone signaling in prostate cancer. Carcinogenesis. 2013

Irizarry RA, Ladd-Acosta C, Wen B, Wu ZJ, Montano C, Onyango P, Cui HM, Gabo K, Rongione M, Webster M, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics. 2009; 41(2):178–186. [PubMed: 19151715]

Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012; 41(1):200–9. [PubMed: 22422453]

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007; 8(1):118–27. [PubMed: 16632515]

Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet. 2002; 3(6): 415–28. [PubMed: 12042769]

Jordan SJ, Whiteman DC, Purdie DM, Green AC, Webb PM. Does smoking increase risk of ovarian cancer? A systematic review. Gynecologic Oncology. 2006; 103(3):1122–1129. [PubMed: 17005245]

Koestler DC, Marsit CJ, Christensen BC, Accomando W, Langevin SM, Houseman EA, Nelson HH, Karagas MR, Wiencke JK, Kelsey KT. Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. Cancer Epidemiol Biomarkers Prev. 2012; 21(8):1293–302. [PubMed: 22714737]

Lehne B, Lewis CM, Schlitt T. From SNPs to Genes: Disease Association at the Gene Level. PLoS One. 2011; 6(6)

Leitzmann MF, Koebnick C, Danforth KN, Brinton LA, Moore SC, Hollenbeck AR, Schatzkin A, Lacey JV. Body Mass Index and Risk of Ovarian Cancer. Cancer. 2009; 115(4):812–822. [PubMed: 19127552]

Leung KC, Doyle N, Ballesteros M, Sjogren K, Watts CK, Low TH, Leong GM, Ross RJ, Ho KK. Estrogen inhibits GH signaling by suppressing GH-induced JAK2 phosphorylation, an effect mediated by SOCS-2. Proc Natl Acad Sci U S A. 2003; 100(3):1016–21. [PubMed: 12552091]

Lu KH, Skates S, Hernandez MA, Bedi D, Bevers T, Leeds L, Moore R, Granai C, Harris S, Newland W, et al. A 2-Stage Ovarian Cancer Screening Strategy Using the Risk of Ovarian Cancer Algorithm (ROCA) Identifies Early-Stage Incident Cancers and Demonstrates High Positive Predictive Value. Cancer. 2013; 119(19):3454–3461. [PubMed: 23983047]

Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008; 454(7203): 436–444. [PubMed: 18650914]

Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, Buckley M, Fridley BL, Tyrer JP, Shen H, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. 2013; 45(4):362–70. 370e1–2. [PubMed: 23535730]

Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SR. The impact of recent alcohol use on genome wide DNA methylation signatures. Front Genet. 2012; 3:54. [PubMed: 22514556]

Rico-Bautista E, Flores-Morales A, Fernandez-Perez L. Suppressor of cytokine signaling (SOCS) 2, a protein with multiple functions. Cytokine Growth Factor Rev. 2006; 17(6):431–9. [PubMed: 17070092]

Romano RA, Li H, Tummala R, Maul R, Sinha S. Identification of Basonuclin2, a DNA-binding zinc-finger protein expressed in germ tissues and skin keratinocytes. Genomics. 2004; 83(5):821–33. [PubMed: 15081112]

Scott M, McCluggage WG, Hillan KJ, Hall PA, Russell SEH. Altered patterns of transcription of the septin gene, SEPT9, in ovarian tumorigenesis. International Journal of Cancer. 2006; 118(5): 1325–1329.

Shen H, Fridley BL, Song H, Lawrenson K, Cunningham JM, Ramus SJ, Cicek MS, Tyrer J, Stram D, Larson MC, et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nat Commun. 2013; 4:1628. [PubMed: 23535649]

Skates SJ, Pauler DK, Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. Journal of the American Statistical Association. 2001; 96(454):429–439.

Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, Anton-Culver H, Chang-Claude J, Cramer DW, DiCioccio R, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet. 2009; 41(9):996–1000. [PubMed: 19648919]

Szabo CI, King MC. Population genetics of BRCA1 and BRCA2. American Journal of Human Genetics. 1997; 60(5):1013–1020. [PubMed: 9150148]

Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One. 2009; 4(12):e8274. [PubMed: 20019873]

Wentzensen N, Black A, Jacobs K, Yang HP, Berg CD, Caporaso N, Peters U, Ragard L, Buys SS, Chanock S, et al. Genetic variation on 9p22 is associated with abnormal ovarian ultrasound results in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. PLoS One. 2011; 6(7):e21731. [PubMed: 21750727]

Whittemore AS, Harris R, Itnyre J. CHARACTERISTICS RELATING TO OVARIAN-CANCER RISK - COLLABORATIVE ANALYSIS OF 12 UNITED-STATES CASE-CONTROL STUDIES .4. THE PATHOGENESIS OF EPITHELIAL OVARIAN-CANCER. American Journal of Epidemiology. 1992; 136(10):1212–1220. [PubMed: 1476143]

Widschwendter M, Apostolidou S, Raum E, Rothenbacher D, Fiegl H, Menon U, Stegmaier C, Jacobs IJ, Brenner H. Epigenotyping in Peripheral Blood Cell DNA and Breast Cancer Risk: A Proof of Principle Study. PLoS One. 2008; 3(7)

Yamanaka T, Matsumoto S, Teramukai S, Ishiwata R, Nagai Y, Fukushima M. The baseline ratio of neutrophils to lymphocytes is associated with patient prognosis in advanced gastric cancer. Oncology. 2007; 73(3-4):215–220. [PubMed: 18424885]

Zhu JG, Dai QS, Han ZD, He HC, Mo RJ, Chen G, Chen YF, Wu YD, Yang SB, Jiang FN, et al. Expression of SOCSs in human prostate cancer and their association in prognosis. Mol Cell Biochem. 2013; 381(1-2):51–9. [PubMed: 23666742]

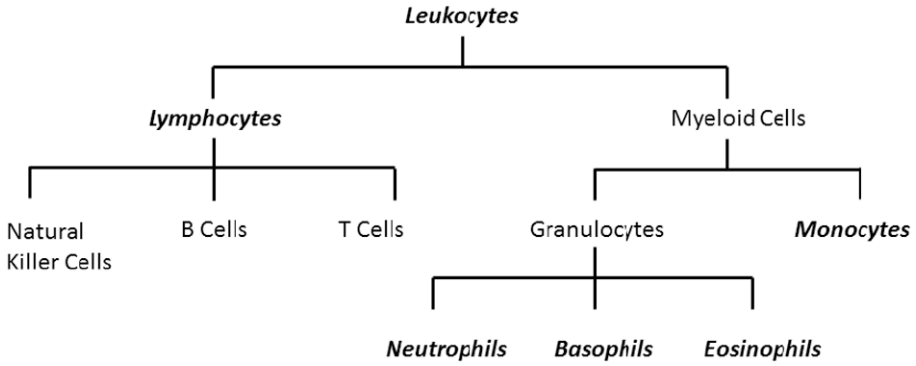

Figure 1.
Diagram of the relationship among CBC measurements.

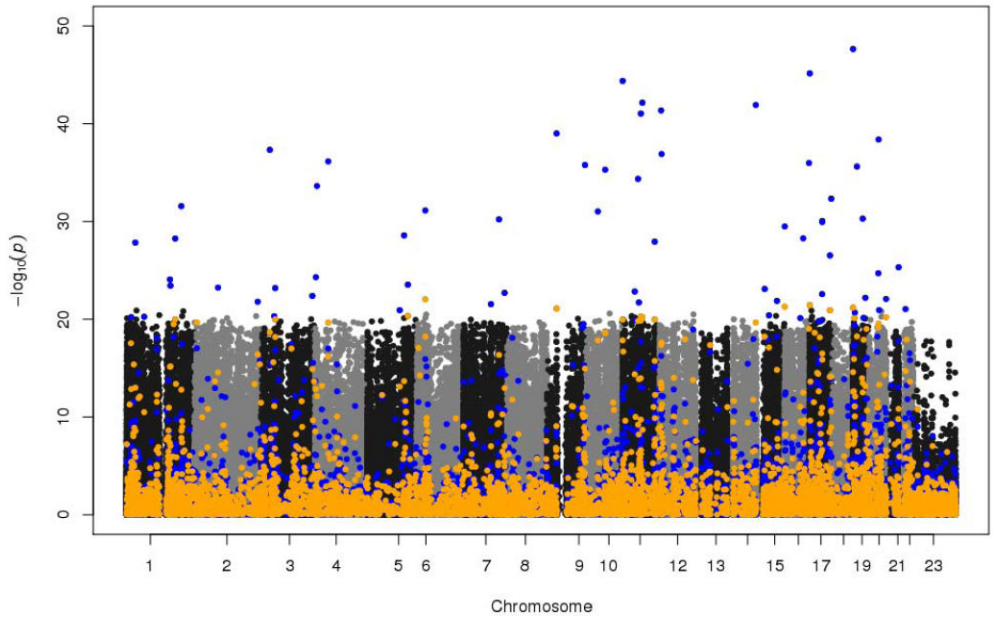

Figure 2.

Manhattan plots of association between CpG probes and case/control status, before CBC adjustment, for probes on the 450K array (black/gray), the 27K array (orange) and the meta-analysis of the probes from the 27K and 450K arrays (blue); other covariates include smoking, alcohol, age at first birth, state, and enrollment year.

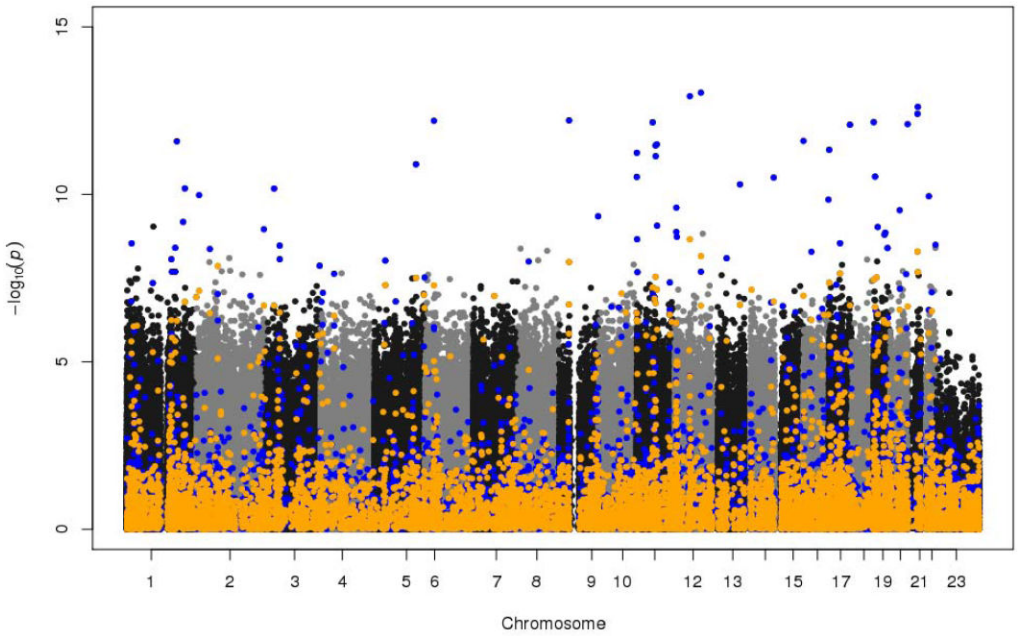

Figure 3.

Manhattan plots of association between CpG probes and case/control status, after CBC adjustment, for probes on the 450K array (black/gray), the 27K array (orange) and the meta-analysis of the probes from the 27K and 450K arrays (blue); other covariates include smoking, alcohol, age at first birth, state, and enrollment year.

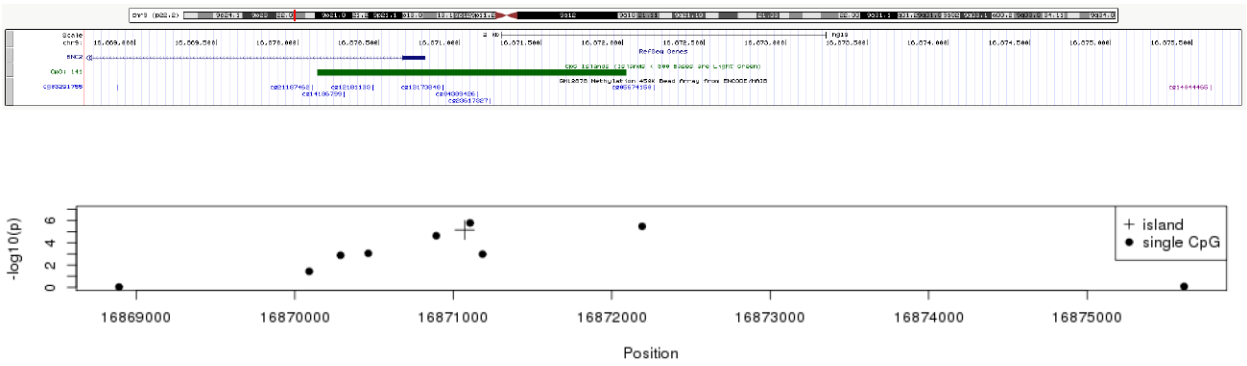

Figure 4.

BNC2 promoter/island region. Top: USCS Genome Browser view of the BNC2 gene promoter and location of CpG probes associated with CpG island 141 (chr9:16870123-16872020). Bottom: -log10(p-values) of case-control association with methylation at each CpG within the BNC2 promoter island region.

Table 1

Clinical characteristics and demographics in the full Mayo sample, and the subsets on the 27K and 450K arrays.

Characteristic	Variable	Level	Total	27K	450K
Clinical	**Case/control, N (%)**	Case	242 (57.2)	153 (58.8)	89 (54.6)
		Control	181 (42.8)	107 (41.2)	74 (45.4)
	Age at diagnosis or enrollment, Mean (SD)	Overall	62.8 (11.3)	62.9 (11.7)	62.7 (10.7)
	Diagnosis	Case	62.8 (11.5)	62.8 (12.1)	62.7 (10.6)
	Enrollment	Control	62.9 (11.0)	63.0 (11.1)	62.7 (11.0)
	Case Histology, N (%)	Low Grade Serous	6 (2.5)	5 (3.3)	4 (4.5)
		High Grade Serous	162 (66.9)	94 (61.4)	33 (37.1)
		Mucinous	7 (2.9)	5 (3.3)	2 (2.2)
		Endometrioid	45 (18.6)	36 (23.5)	9 (10.1)
		Clear Cell	12 (5.0)	10 (6.5)	2 (2.2)
		Other, Epithelial	10 (4.1)	3 (2.0)	39 (43.8)
	Case Grade, N (%)	1	12 (5.0)	8 (5.2)	4 (4.6)
		2	27 (11.3)	22 (14.4)	5 (5.7)
		3	135 (56.3)	80 (52.3)	55 (63.2)
		4	61 (25.4)	43 (28.1)	18 (20.7)
		Unknown	5 (2.1)	0 (0.0)	5 (5.7)
	Case Stage, N (%)	1	35 (14.5)	26 (17.0)	9 (10.1)
		2	12 (5.0)	9 (5.9)	3 (3.4)
		3	157 (64.9)	93 (60.8)	64 (71.9)
		4	38 (15.7)	25 (16.3)	13 (14.6)
		Overall			
Demographic	**Ever Smoking, N (%)**	No	360 (85.1)	229 (88.1)	131 (80.4)
		Yes	35 (8.3)	19 (7.3)	16 (9.8)
		Unknown	28 (6.6)	12 (4.6)	16 (9.8)
		Case			
		No	197 (81.4)	132 (86.3)	65 (73.0)
		Yes	26 (10.7)	15 (9.8)	11 (12.4)
		Unknown	19 (7.9)	6 (3.9)	13 (14.6)
		Control			

Characteristic	Variable	Level	Total	27K	450K
		No	163 (90.1)	97 (90.7)	66 (89.2)
		Yes	9 (5.0)	4 (3.7)	5 (6.8)
		Unknown	9 (5.0)	6 (5.6)	3 (4.1)
		Overall			
	Alcohol, N (%)	Never	109 (25.8)	74 (28.5)	35 (21.5)
		Current	209 (49.4)	124 (47.7)	85 (52.1)
		Former	67 (15.8)	44 (16.9)	23 (14.1)
		Unknown	38 (9.0)	18 (6.9)	20 (12.3)
		Case			
		Never	83 (34.3)	55 (35.9)	28 (31.5)
		Current	99 (40.9)	64 (41.8)	35 (39.3)
		Former	31 (12.8)	22 (14.4)	9 (10.1)
		Unknown	29 (12.0)	12 (7.8)	17 (19.1)
		Control			
		Never	26 (14.4)	19 (17.8)	7 (19.1)
		Current	110 (60.8)	60 (56.1)	50 (67.6)
		Former	36 (19.9)	22 (20.6)	14 (18.9)
		Unknown	9 (5.0)	6 (5.6)	3 (4.1)
		Overall			
	Age at first birth, N (%)	Nulliparous	60 (14.2)	36 (13.8)	24 (14.7)
		<=20 yrs	99 (23.4)	59 (22.7)	40 (24.5)
		>20 yrs	247 (58.4)	157 (60.4)	90 (55.2)
		Unknown	17 (4.0)	8 (3.1)	9 (5.5)
		Case			
		Nulliparous	38 (15.7)	24 (15.7)	14 (15.7)
		<=20 yrs	67 (27.7)	41 (26.8)	26 (29.2)
		>20 yrs	129 (53.3)	86 (56.2)	43 (48.3)
		Unknown	8 (3.3)	2 (1.3)	6 (6.7)
		Control			
		Nulliparous	22 (12.2)	12 (11.2)	10 (13.5)
		<=20 yrs	32 (17.7)	18 (16.8)	14 (18.9)

Characteristic	Variable	Level	Total	27K	450K
		>20 yrs	118 (65.2)	71 (66.4)	47 (63.5)
		Unknown	9 (5.0)	6 (5.6)	3 (4.1)
CBC	**Leukocytes, Mean (SD)**	Overall	7.74 (3.19)	7.87 (3.25)	7.54 (3.10)
		Case	9.02 (3.48)	9.05 (3.53)	8.97 (3.40)
		Control	6.04 (1.61)	6.19 (1.73)	5.83 (1.39)
	Neutrophils, Mean (SD)	Overall	5.57 (3.13)	5.66 (3.18)	5.42 (3.06)
		Case	6.97 (3.36)	6.93(3.44)	7.04 (3.24)
		Control	3.69 (1.30)	3.84 (1.44)	3.47 (1.04)
	Lymphocytes, Mean (SD)	Overall	1.44 (0.60)	1.47 (0.60)	1.38 (0.61)
		Case	1.24 (0.55)	1.31 (0.57)	1.12 (0.50)
		Control	1.70 (0.56)	1.69 (0.55)	1.70 (0.57)
	Monocytes Mean, (SD)	Overall	0.57 (0.28)	0.57 (0.27)	0.57 (0.28)
		Case	0.65 (0.32)	0.64 (0.32)	0.67 (0.33)
		Control	0.46 (0.15)	0.46 (0.14)	0.45 (0.15)
	Eosinophils, Mean (SD)	Overall	0.14 (0.14)	0.13 (0.13)	0.14 (0.16)
		Case	0.12 (0.15)	0.12 (0.14)	0.12 (0.16)
		Control	0.15 (0.13)	0.15 (0.11)	0.16 (0.15)
	Basophils, Mean (SD)	Overall	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
		Case	0.03 (0.02)	0.03 (0.03)	0.03 (0.02)
		Control	0.04 (0.02)	0.03 (0.02)	0.04 (0.02)
	Proportion of Neutrophils, Mean (SD)	Overall	0.69 (0.12)	0.69 (0.12)	0.69 (0.12)
		Case	0.75 (0.09)	0.75 (0.10)	0.77 (0.09)
		Control	0.60 (0.09)	0.61 (0.10)	0.59 (0.08)

Table 2

Overlapping CpGs on 27K and 450K (N=22,278 probes), replicating association between methylation and case/control status (meta-analysis p<5E-8). Analyses are adjusted for proportion of neutrophils, age at first birth, current smoking, current alcohol consumption, state of residence, and enrollment year.

				27K Array		450K Array		Meta-Analysis*	
CpG Probe	**Chr**	**Gene**	**CpG Island**	**Effect**	**P-value**	**Effect**	**P-value**	**Effect**	**P-value**
cg11738543	12	*SOCS2*	Island	-0.043	6.96E-09	-0.050	2.03E-05	-0.045	9.17E-14
cg23140706	12	*NFE2*		-0.038	2.20E-09	-0.047	7.87E-05	-0.040	1.17E-13
cg03801286	21	*KCNE1*		-0.060	2.10E-08	-0.078	1.25E-05	-0.065	2.45E-13
cg17356733	21	*IFNGR2*	Shore	-0.061	5.10E-09	-0.082	7.97E-05	-0.065	3.93E-13
cg15248035	9	*CCIN*	Shelf	-0.054	1.04E-08	-0.073	5.54E-05	-0.058	6.20E-13
cg24211388	6	*AIF1*		-0.062	5.19E-08	-0.078	1.26E-05	-0.066	6.35E-13
cg18084554	19	*ARID3A*	Shore	-0.068	3.52E-08	-0.070	2.73E-05	-0.069	6.97E-13
cg23547429	11	*SLC43A3*	Shore	-0.052	5.99E-08	-0.057	1.55E-05	-0.054	7.06E-13
cg09868035	20	*C20orf135*	Shore	-0.054	2.22E-07	-0.072	3.21E-06	-0.060	8.02E-13
cg03330678	17	*SEPT9*		-0.060	2.07E-07	-0.072	4.49E-06	-0.064	8.33E-13
cg16003913	16	*MPG*	Shore	-0.046	1.06E-07	-0.058	2.13E-05	-0.049	2.54E-12
cg19399532	1	*C1orf220*	Shore	-0.049	5.98E-07	-0.060	4.30E-06	-0.053	2.60E-12
cg25634666	11	*FOLR3*		-0.055	6.91E-08	-0.071	4.07E-05	-0.059	3.21E-12
cg22224704	11	*GSTP1*	Shore	-0.054	2.89E-08	-0.055	1.45E-04	-0.055	3.46E-12
cg22242539	17	*SERPINF1*		-0.052	4.36E-08	-0.055	1.21E-04	-0.053	4.67E-12
cg22016649	11	*PNPLA2*	Shore	-0.044	3.52E-07	-0.054	1.41E-05	-0.047	5.74E-12
cg07730301	11	*ALDH3B1*		-0.055	1.75E-07	-0.072	3.03E-05	-0.059	7.26E-12
cg01965939	5	*SH3TC2*		-0.038	3.12E-08	-0.040	4.45E-04	-0.039	1.26E-11
cg09001777	19	*FUT3*		-0.045	2.34E-07	-0.054	1.05E-04	-0.048	2.95E-11
cg00645579	11	*IRF7*	Shore	-0.056	5.66E-07	-0.076	1.02E-06	-0.064	3.01E-11
cg23090046	14	*KLC1*	Island	-0.055	1.60E-07	-0.052	1.84E-04	-0.054	3.15E-11
cg10057295	13	*STK24*	Shore	-0.042	2.00E-07	-0.057	1.54E-04	-0.045	5.02E-11
cg12380764	1	*IL19*		-0.047	1.62E-07	-0.063	2.60E-04	-0.050	6.65E-11
cg01617750	3	*CMTM8*	Shore	-0.054	2.10E-07	-0.045	2.45E-04	-0.050	6.71E-11

				27K Array		450K Array		Meta-Analysis*	
CpG Probe	**Chr**	**Gene**	**CpG Island**	**Effect**	**P-value**	**Effect**	**P-value**	**Effect**	**P-value**
cg11827101	2	*LOC339789*		-0.032	7.59E-08	-0.037	1.20E-03	-0.033	1.05E-10
cg04541607	22	*CRYBB1*		-0.035	1.38E-06	-0.048	4.16E-05	-0.038	1.14E-10
cg13745346	16	*CBFA2T3*		-0.040	3.89E-07	-0.048	2.67E-04	-0.042	1.44E-10
cg15784615	12	*LTBR*	Shore	-0.046	1.23E-05	-0.053	1.48E-05	-0.049	2.50E-10
cg23713742	20	*SPAG4*	Shore	-0.041	2.77E-07	-0.042	8.71E-04	-0.041	2.97E-10
cg08529852	9	*EGFL7*	Shore	-0.046	1.02E-05	-0.047	3.35E-05	-0.047	4.52E-10
cg21870884	1	*GPR25*	Island	-0.040	1.21E-04	-0.051	3.59E-06	-0.045	6.64E-10
cg10287137	11	*P2RY2*	Island	-0.055	1.09E-06	-0.050	5.47E-04	-0.053	8.63E-10
cg00666746	19	*SYDE1*	Shore	-0.036	4.86E-06	-0.048	9.78E-05	-0.040	9.44E-10
cg09039163	2	*ESPNL*	Shore	-0.054	2.10E-07	-0.048	4.12E-03	-0.052	1.10E-09
cg27347104	12	*VWF*		-0.030	6.88E-07	-0.037	1.19E-03	-0.032	1.32E-09
cg21126943	19	*CEACAM6*		-0.042	1.85E-06	-0.057	3.71E-04	-0.045	1.40E-09
cg06394229	19	*LGALS4*		-0.032	6.76E-07	-0.033	1.70E-03	-0.033	1.59E-09
cg02082571	12	*CLEC4A*		-0.035	9.81E-06	-0.048	8.62E-05	-0.039	1.85E-09
cg25623459	11	*TNNI2*		-0.032	6.21E-07	-0.038	2.25E-03	-0.033	2.19E-09
cg13053608	17	*GHDC*		-0.046	2.30E-08	-0.066	4.54E-05	-0.052	2.88E-09
cg18333690	1	*PADI4*		-0.037	5.67E-06	-0.051	2.19E-04	-0.041	2.91E-09
cg04586023	22	*HDAC10*	Shore	-0.038	8.58E-06	-0.054	6.41E-05	-0.043	3.21E-09
cg17386185	3	*GLYCTK*	Shore	-0.026	2.03E-06	-0.034	8.79E-04	-0.028	3.40E-09
cg10161121	1	*FASLG*		0.037	5.30E-06	0.046	3.92E-04	0.040	3.94E-09
cg15528736	19	*FCGRT*	Shore	-0.054	7.20E-07	-0.077	8.66E-06	-0.062	3.99E-09
cg27485921	2	*ATP6V1E2*		-0.046	3.53E-07	-0.066	4.20E-05	-0.052	4.27E-09
cg25341726	16	*IL27*		-0.034	5.66E-07	-0.053	4.09E-03	-0.036	5.18E-09
cg24821554	13	*GUCY1B2*		-0.044	2.36E-06	-0.049	2.10E-03	-0.045	8.08E-09
cg10061138	3	*STAB1*		-0.044	3.33E-07	-0.062	7.81E-07	-0.051	8.60E-09
cg10636246	1	*AIM2*		-0.031	5.69E-07	-0.044	8.27E-03	-0.032	8.62E-09
cg27606341	5	*FYB*		-0.054	5.13E-08	-0.077	3.14E-05	-0.061	9.49E-09
cg27258399	8	*HTRA4*	Island	-0.048	1.74E-05	-0.069	2.13E-04	-0.054	1.01E-08

				27K Array		450K Array		Meta-Analysis*	
CpG Probe	**Chr**	**Gene**	**CpG Island**	**Effect**	**P-value**	**Effect**	**P-value**	**Effect**	**P-value**
cg27033479	3	*TMEM44*	Shore	-0.024	2.22E-05	-0.035	2.27E-04	-0.027	1.35E-08
cg00071250	1	*FASLG*		0.042	2.39E-06	0.037	5.05E-03	0.041	2.04E-08
cg17166812	1	*NDUFS2*	Shelf	-0.031	1.88E-06	-0.032	6.99E-03	-0.031	2.04E-08
cg23412850	12	*SOCS2*	Island	-0.034	2.12E-07	-0.049	3.90E-05	-0.039	2.05E-08
cg21019522	11	*SLC22A18*	Shelf	-0.041	2.29E-05	-0.057	3.52E-04	-0.045	2.10E-08
cg02600394	4	*TXK*		0.040	4.16E-07	0.027	3.11E-02	0.037	2.34E-08
cg23265096	20	*CTSZ*	Island	-0.044	6.02E-06	-0.046	2.32E-03	-0.044	2.41E-08
cg11394785	5	*LTC4S*	Shore	-0.032	1.47E-05	-0.046	1.95E-04	-0.036	3.04E-08
cg07728874	11	*CD3D*		0.030	1.54E-06	0.030	1.89E-02	0.030	4.27E-08
cg20125091	1	*GFI1*	Shore	-0.035	5.15E-06	-0.040	4.75E-03	-0.036	4.44E-08

*Batch heterogeneity was used using Cochrane's Q statistic. For all probes, p>0.2.

Table 3

PCA results of R=25,607 CpG island regions in the 450K data, p<1E-5.

Chr	Island/Shore/Shelf*	Nearest Gene	Near Promoter Region?	Number of Probes per Region	% Variation Explained	P-value
9	chr9:16870123-16872020	*BNC2*	Yes	9	40.9	7.20E-06
4	chr4:57975860-57976916	*IGFBP7*	Yes	15	48.2	7.57E-06
18	chr18:77794360-77794761	*C18orf22*	Yes	9	50.5	8.61E-06
7	chr7:152372980-152373421	*XRCC2*	Yes	13	35.6	8.62E-06
12	chr12:133067072-133067735	*FBRSL1*	Yes	19	32.1	8.81E-06
10	chr10:8091374-8098329	*GATA3/GATA3-AS1*	Yes	73	35.2	9.80E-06

*Designations of CpG islands, shores, and shelves are based on Genome Build 37.