A Methylation Density Binary Classifier for Predicting and Optimizing the Performance of Methylation Biomarkers in Clinical Samples

Brendan F. Miller[1], Thomas R. Pisanic II*[,2], Gennady Margolin[1], Hanna M. Petrykowska[1], Pornpat Athamanolap[5], Akosua Osei-Tutu[3], Christina M. Annunziata[3], Tza-Huei Wang[2,4,5], Laura Elnitski*[,1]

[1] Translational Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, 20892, USA

[2] Institute for NanoBioTechnology, Johns Hopkins University, Baltimore, Maryland, 21218, USA

[3] Women's Malignancy Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, USA

[4] Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland, 21218, USA

[5] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, 21218, USA

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

* To whom correspondence should be addressed. Tel: +1 301 451 0265; Fax: +1 301 435 6170; Email: elnitski@mail.nih.gov. Correspondence may also be addressed to Thomas R. Pisanic II. Tel: +1 619 892 2567; Fax: +1 410 516 2355; Email: tpisanic@jhu.edu.

Present Address: Laura Elnitski, Translational Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, 20892, USA; Thomas R. Pisanic II, Institute for NanoBioTechnology, Johns Hopkins University, Baltimore, Maryland, 21218, USA.

ABSTRACT

Aberrant DNA methylation is commonly heralded as a promising cancer biomarker; however, its inherently stochastic nature often leads to variable methylation patterns that can complicate the use of methylation

biomarkers for clinical diagnostics, particularly in dilute samples such as liquid biopsies. Here, we present a methylation density binary classifier, a statistical method for leveraging differential heterogeneous methylation to predict and optimize the performance of methylation biomarkers for clinical applications. We first developed and tested the classifier using methylation density profiles derived from reduced representation bisulfite sequencing reads of ovarian carcinoma at *ZNF154*, a recurrently methylated locus in multiple cancer types. We then used *in silico* simulations to predict the performance of the classifier in liquid biopsies and validated these predictions using quasi-digital melt curve analysis (DREAMing) of circulating cell-free DNA from individuals with versus without ovarian carcinoma. We found good agreement between predicted and observed classifier performance, and further demonstrated that implementation of this approach with *ZNF154* outperformed CA-125 for use in etiologically-diverse ovarian cancer types. Our results indicate that methylation density profiles can be exploited to predict and facilitate implementation of methylation biomarkers for clinical applications, and that *ZNF154* methylation shows promise as a clinically-useful biomarker for ovarian cancer.

INTRODUCTION

A primary aim in cancer diagnostics is to identify and develop reliable biomarkers for malignancy detection that can ideally be detected with non-invasive detection methods. Epigenetic biomarkers are potentially useful cancer indicators, as many genomic loci become recurrently methylated during tumorigenesis and can thus serve as reliable indicators of disease (1-3). A particularly attractive application for methylation biomarkers is use in liquid biopsies, which represent minimally invasive assays to detect circulating tumor DNA (ctDNA), and associated epigenomic alterations, when shed from cancerous tissues into the blood (4). While whole methylome analyses, such as whole-genome bisulfite sequencing (WGBS) (5) and Infinium BeadArrays (6), have identified scores of differentially-methylated genomic loci in cancers (7), only a handful of individual loci have ultimately become effective for use as biomarkers in clinical applications (8,9). This is due to a number of technical hurdles involved in translating promising methylation biomarkers for use in liquid biopsies, including (i) the small proportion of plasma ctDNA relative to cell-free DNA (cfDNA) derived from healthy cells (10), (ii) heterogeneity of methylation at a given locus (11-14), (iii) age-associated accrual of methylation confounding marker selection (15), and (iv) differences in the yield of extracted

cfDNA between samples (16). Collectively, these issues often make it difficult to achieve the high degrees of sensitivity and specificity necessary to attain reasonable clinical performance with a limited number of biomarkers (17). Thus, there remains a need for the development and implementation of new methods capable of increasing the signal to noise ratio in order to harness the diagnostic potential of methylated biomarkers in general.

The overall aim of the present study was to investigate, and potentially exploit, differences in locus-specific DNA methylation heterogeneity between case and control samples to optimize the use of methylation biomarkers in blood-based detection methods. We introduce a methylation density binary classifier (MDBC) aimed at maximizing the performance of methylation biomarkers by leveraging statistical differences in the profiles of epiallelic methylation densities between sample cohorts. Methylation density, or the proportion of methylated CpG-dinucleotides in a given genomic locus, leverages the degree of CpG methylation to overcome background methylation noise attributed to technical or biological conditions (8,18,19). The measurement is agnostic to changes within the methylation pattern and therefore has the potential to outperform methods that are reliant on methylated positions such as methyl specific PCR, and to achieve high sensitivity. Specifically, we examine methylation at the *ZNF154* locus, which we previously demonstrated to be differentially-methylated in at least 14 different solid cancer types (2,20). More recently, reduced representation bisulfite sequencing (RRBS) of epithelial ovarian carcinomas (EOCs) also identified the *ZNF154* locus among the most promising biomarker candidates in tissue and serum specimens (8).

In the present work, we employ this RRBS data to develop a framework for the MDBC as well as to estimate the potential utility of the classifier to improve *ZNF154* diagnostic performance in dilute samples such as liquid biopsies. We validate the MDBC in liquid biopsy specimens using our previously-reported, ultra-sensitive high-resolution melting assay, Discrimination of Rare EpiAlleles by Melt (DREAMing), which provides an inexpensive and straightforward means of evaluating locus-specific methylation density at single-molecule sensitivity and single CpG-site resolution (21). We employ DREAMing to assess the methylation density profiles of plasma samples obtained from 34 patients with refractory EOC and 57 cancer-free controls. From this data we demonstrate significant improvement in the ability to distinguish between patient and controls plasma samples using MBDC compared to mean locus methylation signal. Lastly, we show that our method more accurately identifies EOC in liquid biopsies from etiologically diverse

tumor subtypes than measurement of blood CA-125, the most commonly-employed biomarker for monitoring EOC. Not only do these results suggest potential as an improved screening method for ovarian cancer, they more broadly offer a practical means of improving the diagnostic performance of DNA methylation-based biomarkers, particularly in challenging samples such as liquid biopsies.

MATERIAL AND METHODS

Datasets and samples

450K Illumina Infinium HumanMethylation450 BeadChip datasets. Processed Illumina Human 450K data for EOCs (GEO accession GSE72021, described in (8)) from 221 tumor samples (171 serous, 18 endometrioid, 14 clear cell, 9 mucinous and 9 other histological cancer subtypes) and WBCs (GEO accession GSE55763, described in (22)) from 2,664 individuals was downloaded from the NCBI's Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). Data for TCGA solid tumor and control sample sets was downloaded from the Broad Institute (https://gdac.broadinstitute.org/) FireHose. Beta-values for probes cg11294513, cg05661282, cg21790626, and cg27049766 were extracted using custom Python-2.7.14 (https://python.org) scripts.

Reduced representation bisulfite sequencing data. Reduced representation bisulfite sequencing (RRBS) data was obtained for 12 EOCs, 10 healthy ovarian tissues, and 22 WBC samples used in Widschwendter *et al.* (8). The data was downloaded from the European Genome-phenome Archive (dataset accession: EGAD00001003822).

Plasma samples. Plasma samples (n=107) were obtained from 34 patients with late-stage residual EOC and 57 pathologically normal control patients. The patients with EOC were part of two different clinical trials, described below, and controls were recruited by Fox Chase Cancer Center. Samples were split into two cohorts. One was used to assess the performance of the MDBC and establish optimal cutoffs in plasma and the second cohort was used to validate the MDBC cutoffs in plasma and compare the cutoffs to CA-125.

The first cfDNA cohort contained plasma from 26 EOC-positive patients and 45 healthy women (1 sample per patient). The patients with EOC had previously received standard-of-care treatment (a platinum- and taxane-containing regimen), relapsed after one or more subsequent treatment regimens, and been recruited into a clinical trial testing combined treatment with bevacizumab and dasatinib (NCT01445509). The plasma samples were taken at baseline before treatment was administered. Volumes of plasma processed and extracted cfDNA concentrations can be found in Supplemental Table 1.

The second cohort encompassed 24 plasma samples from 8 patients with ovarian cancer (3 samples per patient) and 12 control samples from healthy women (1 sample per patient). Here, too, patients with ovarian cancer were given standard of care, relapsed, and were recruited to a clinical trial, this one testing combined treatment with bevacizumab and sorafenib (NCT00436215). Ovarian cancer patient plasma samples were taken at three separate time points over 6 weeks (at baseline, 2 weeks after exposure to the first agent, and 2 weeks after exposure to the second agent). Blood was collected into standard EDTA tubes and maintained on ice for transport; plasma was separated immediately using centrifugation and frozen in 1.0- to 1.5-ml aliquots at -80°C. Additionally, CA-125 protein levels in the blood were measured every 4 weeks using standard immunoassay testing.

Mean locus methylation and weighted sample fraction of methylation density at the *ZNF154* locus from RRBS data

The RRBS data were aligned to hg19 using Bismark-0.19.0 (23). Counts of RRBS reads that overlapped the region Chr19:58220000-58220800 were tallied for each sample and divided into subgroups based on their specific start coordinates and combinations of methylated and unmethylated cytosines. Based on sample metadata, counts of reads from replicate libraries of the same sample were pooled together. The three largest read pileups, which encompassed approximately 95% of the total reads, were found to have start coordinates located at chr19:58220394, chr19: 58220395, chr19: 58220396, chr19: 58220535, chr19: 58220536, chr19: 58220537, chr19: 58220122, chr19: 58220123, and chr19: 58220124. Reads not containing these start coordinates were ignored. One WBC sample did not have any reads at these coordinates and therefore was removed from the analysis.

Methylation density was defined as the number of methylated CpGs (meCpGs) out of total CpGs in a given read or DNA molecule. The weighted methylation level of a locus, or henceforth the mean locus methylation, was defined as the total number of meCpGs out of all CpGs sequenced in reads from the locus in question (24). The sample epiallelic fraction was defined as the proportion of reads with a given methylation density out of all reads from a given locus.

Reads with the same methylation density but having different numbers of CpGs were weighed in terms of their contribution to the overall sample methylation density by normalizing their epiallelic fractions in terms of the number of CpGs they covered out of the total CpGs covered in all of the reads. For this, we calculated the epiallelic fraction for a given set of reads by:

$$\frac{C}{Ct}$$

where C = # of CpGs covered by the reads with a methylation density above a given cutoff and Ct = total CpGs in all reads.

Methylation density binary classifier (MDBC) procedure

The DREAMing assay includes two parameters that must be specified before a sample can be called positive: 1) a minimum methylation density MD_{min}, and 2) a minimum epiallelic fraction EF_{min}. A sample is considered positive if at least EF_{min} of the DNA fragments are methylated at a density above or equal to MD_{min}, where the optimal parameters are chosen to maximize sensitivity+specificity (or, equivalently, true positive rate (TPR) – false positive rate (FPR)). Given a set of training data, the MDBC procedure solves for these two parameters simultaneously by calculating TPR-FPR for various combinations of MD_{min} and EF_{min}, choosing the pair of parameter values that maximizes the value.

In practice we define a range of values for MD_{min}, from 0 to 1 in increments of 0.05, and let the data define an appropriate range of EF_{min} for each possible density MD_i by considering the full set of epiallelic fractions that are observed in the training data, at the selected MD_i. In the extreme case that MD_{min}=0, we consider fragments that have a density above but not equal to MD_{min} so that we are considering all except fully unmethylated reads for computing the sample EF.

For the RRBS samples, the epiallelic fractions are the weighted sample fractions described above. For the plasma samples methylation density is derived from the melting temperatures as described below,

and epiallelic fractions are calculated from the DREAMing melt peaks representative of each DNA divided by the total genomic equivalents assessed.

Simulated dilution and classification performance comparison between the mean locus methylation and MDBC using RRBS data

To test the classification performance of either using mean locus methylation or the optimal methylation density and sample epiallelic fraction determined by the MDBC across a range of tumor signals, RRBS reads from EOCs or WBCs were randomly mixed together at different fractions at a set depth of total reads. A total depth of 10,000 reads was determined in order to simulate dilutions down to 0.01%. At each dilution, a simulated set of "spike-in" samples was made by randomly pairing each of the 12 EOCs with one of the 22 WBC samples for a total spike-in sample set size of 22. For each pair, a distribution of read methylation densities was generated for the EOC and the WBC sample based on the weighted sample fractions of methylation densities. Reads were randomly sampled from each distribution at a ratio equivalent to the % tumor dilution in question with the total number of reads sampled equal to the total read depth. For each dilution, receiver operating characteristic (ROC) curves were built for classifying samples using either the mean locus methylation level or the sample epiallelic fraction based on the methylation density cutoffs selected by the MDBC. The spike-ins were used as cases and the original 22 WBC samples as controls. ROC area under the curve (AUC) for the mean locus methylation and the optimal methylation density were recorded. This simulation was repeated 50 times and the mean AUC and the 95% CI for the mean locus methylation and MDBC methylation density were computed.

Using the same 50 iterations of the simulated dilution mentioned above, we compared the performance of using the mean locus methylation or a methylation density cutoff at high dilutions of EOC RRBS reads. For select tumor dilutions (1.0%, 0.1%, 0.01%), we recorded AUC values for each methylation density cutoff as well as the sensitivity and specificity resulting from the optimal epiallelic fraction cutoff for that methylation density. Using the AUCs, we then determined the probability of achieving an improved performance over mean locus methylation for each methylation density cutoff. Given that the AUC can be defined as: "the probability that a randomly selected case will have a higher test result than a randomly selected control," (25) we considered each application of the methylation density cutoff a case and the

resulting AUC its test value whereas the mean locus methylation AUCs would be the test results for a set of controls. We constructed a new ROC based on these two sets of AUC values (n=50 for both methylation density and mean locus methylation) and the resulting AUC of this ROC was considered the probability that the methylation density cutoff would produce a higher AUC, or improved performance, than the mean locus methylation.

Measurement of ZNF154 methylation using DREAMing

Plasma cfDNA extraction and bisulfite conversion. To extract cfDNA, we employed the methylation-on-beads protocol (26) using NeoGeneStar Cell Free DNA Purification kits with pretreatment reagents and NeoGeneStar magnetic beads (NeoGeneStar, Somerset, NJ). Extracted cfDNA was bisulfite converted using Zymo Lightning Conversion reagents (Zymo Research, Irvine, CA) and eluted twice in 50 μl Zymo Elution Buffer using 1.5 ml LoBind Eppendorf tubes (Eppendorf, Hauppage, NY). We quantified the resulting DNA by performing qPCR in duplicate, using a primer and a TaqMan probe set to amplify a 100-bp region overlapping the bisulfite-converted top strand of the beta-actin locus on a C1000 Touch Thermo Cycler (BioRad, Hercules, CA) using a CFX96 Real-Time System.

Quantifying ZNF154 methylation in cfDNA with DREAMing. The DREAMing method quantifies methylation density via melting-peak analysis of bisulfite-treated sample DNA (21). Bisulfite conversion deaminates unmethylated cytosines to uracils, which eventually are replaced by thymines during amplification. By contrast, methylated cytosines remain intact. After PCR amplification, the difference in the base-stacking energy of C/G vs. A/T bases creates distinct melting temperatures in the products; more C/G pairs result in a higher melting temperature. DREAMing improves upon previous methylation-sensitive high-resolution melting assays (27,28) by diluting the sample to achieve quasi-digitization of the population of methylated fragments, such that each reaction contains target DNA fragments with no more than two different methylation densities (i.e., epialleles). This helps eliminate the formation of heteroduplexes, allowing us to determine the methylation density of the respective epialleles.

We used DREAMing to calculate the fraction of *ZNF154* DNA fragments with a given methylation density. First, we determined that the maximum number of genomic copies of cfDNA that could be loaded

into a well without compromising the detection of single epialleles of the targeted *ZNF154* genomic region was approximately 400, based on our ability to detect single fragments of fully methylated synthetic *ZNF154* target in a known quantity of low-methylation bisulfite-treated genomic DNA. We partitioned each sample across 12 wells on a 96-well microtiter plate, for a total of 8 samples per plate. Thus, the maximum amount of a given sample that we could query in a single DREAMing assay would be 4,800 genomic copies (12 wells times 400 genomic copies per well). Each sample was queried via DREAMing at least twice, for a total of 24 wells per sample, and the data for multiple runs of a given sample were pooled together. The fraction of methylated *ZNF154* genomic fragments for each sample was determined by counting the number of melting peaks above a defined temperature cutoff (corresponding to a specific methylation density cutoff, or number of methylated CpGs out of 14 total positions for our *ZNF154* locus; for conversion between these two values see next section), where each counted peak corresponded to a single methylated cfDNA fragment, inferred from Poissonian statistics. We divided these counts by the total number of genomic copies loaded into the DREAMing assay (representative of the total number of cfDNA fragments analyzed) to arrive at the sample epiallelic fraction of cfDNA fragments with a given methylation density.

The reaction conditions for DREAMing were as follows: Master PCR mixes were made so that each well would have a final volume of 25 μl with 200 μM dNTP mixture, 300 nM forward *ZNF154* DREAMing primer, 300 nM reverse *ZNF154* DREAMing primer, 1X EvaGreen (Biotium Inc, Fremont, CA), 0.04 U/μl Platinum Taq (Thermo Fisher Scientific, Bothell, WA), and 1X in-house “Magic Buffer” (16.6 mM ammonium sulfate, 67 mM Tris pH 8.8, 2.7 mM magnesium chloride, and 10 mM beta-mercaptoethanol). DREAMing reactions were run on a C1000 Touch Thermal Cycler using a CFX96 Real-Time System. Reactions were run at 95°C for 5 minutes for 1 cycle; 95°C for 30 seconds, 61.4°C for 30 seconds, and 72°C for 30 seconds for 50 cycles; followed by a temperature gradient beginning at 65°C and ramping up to 90°C in 0.2°C increments, each held for 10 seconds, before SYBR/FAM fluorescence was imaged. After DREAMing, melting temperature peaks were visualized using the accompanying CFX Manager 3.1 software to analyze the negative derivative of the change in fluorescence (-d(RFU)/dT) versus temperature plots for each well.

Validation and calibration of DREAMing melt peak temperatures to methylation densities. Because accurate determination of methylation density depends on efficient bisulfite conversion, we used bisulfite amplicon

sequencing to validate the results of at least one well per sample in the training set from the DREAMing assay. We focused on wells that contained a temperature peak indicative of a high methylation density cfDNA fragment and when possible, selected multiple wells that together were representative of the overall methylation profile for a given sample. In this way we included wells that had cfDNA fragments with low methylation densities and wells that had high methylation densities and sought to test the validity of the assay across this methylation range.

For each selected DREAMing well, 20 μl was pipetted into a new well on a 96-well plate and cleaned using Ampure XP beads (Beckman Coulter, Brea, CA) according to the manufacturer's protocol, at a ratio of 1.8 μl beads to 1 μl sample. The DNA was then eluted with 35 μl EB buffer (Qiagen, Germantown, MD), and 30.75 μl of the elution was combined with 5 μl 10X TaKaRa EpiTaq PCR Buffer (Mg^{2+} free; TaKaRa, Mountain View, CA), 5 μl 25 mM $MgCl_2$, 6 μl 2.5 mM dNTP mixture, 1 μl 12.5 μM forward primer (175-bp forward *ZNF154* DREAMing primer with sequencing adapter), 1 μl 12.5 μM reverse primer (175-bp reverse *ZNF154* DREAMing primer with sequencing adapter), 1 μl DMSO, and 0.25 μl of 5 U/μl TaKaRa EpiTaq DNA polymerase, for a total reaction volume of 50 μl. This mixture was placed in a SimpliAmp thermal cycler (Applied Biosystems, Foster City, CA) using the following conditions: 95°C for five minutes and one cycle; 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds, for nine cycles; and 72°C for seven minutes for one cycle. We then performed a second Ampure XP beads cleanup by combining 46 μl each PCR reaction with 55 μl beads, eluting in 27 μl EB buffer. Next, 23 μl elution was combined with 25 μl 2X High Fidelity Phusion Master Mix (NEB, Ipswich, MA), and 1 μl i7 and i5 barcoding primers for a total reaction volume of 50 μl. Another round of PCR was performed under the following conditions: 98°C for 30 seconds and one cycle; 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 30 seconds, for nine cycles; and 72°C for five minutes and one cycle. After this, each reaction was cleaned again with Ampure XP beads using 55 μl beads and 46 μl sample and eluted in 30 μl EB buffer. Then 3 μl elution was run on a 2% agarose gel to confirm the expected band of 300-bp (size of amplicon with adapter and barcodes). Samples were submitted to the NIH Intramural Sequencing Center for quality control and sequencing on a MiSeq using 300-bp paired end sequencing. Using Bismark-0.19.0, analysis of reads, bisulfite conversion efficiency, and determination of meCpG patterns was performed as described

previously (20). Wells with sequenced amplicons that had less than 95% bisulfite conversion efficiency were discarded.

We implemented a linear regression model to convert DREAMing melt peak temperatures to methylation density values determined by the bisulfite amplicon sequencing. Patterns of meCpGs computed from the sequencing reads were kept for a given well if their abundance was above 1%. Sequencing patterns were ordered based on methylation density and abundance and the pattern with the highest methylation density from this list was then matched to the highest melt peak temperature. Given that our unmethylated controls consistently melted between 80°C and 80.4°C, we also discarded melt temperatures below 80°C for this calibration. We found good agreement between the DREAMing melt peak temperature and the actual methylation density determined from the bisulfite amplicon sequencing (Fig S3). We used the generated linear model to convert all melt peak temperatures into methylation densities and rounded these to the nearest 7% given that we would expect each additional meCpG in our locus of 14 potential meCpG sites to increase the methylation density by 1/14 or approximately 7%.

Mean locus methylation in plasma. The epiallelic fraction of all detected epialleles via DREAMing except fully unmethylated DNA fragments (based on a methylation density cutoff of 0%) was used as an estimate of the sample mean locus methylation. This was based on the fact that, unlike the RRBS reads, the number of CpG sites covered (14) was the same for each DNA fragment targeted in the DREAMing assay. Therefore, the fraction of these CpGs that were methylated would be proportional to the fraction of total methylated epialleles.

Statistical analyses and plotting

Plotting and statistical analyses were performed using custom Python-2.7.14 (https://python.org) and R-3.4.4 (29) scripts. To compare epiallelic fractions or mean locus methylation between groups, boxplot comparisons were performed. Statistical significance was evaluated using the two-sided Wilcoxon rank sum test. The true and false positive rates, associated thresholds, as well as the area under the curve (AUC) for the receiver operating characteristic analyses were generated using the python package sklearn (v0.19.1) (30) with the module sklearn.metrics. AUC 95% confidence intervals were computed using the R

library pROC (31) and using the ci.auc() command with method="bootstrap" on the data using 2000 stratified replicates. Statistical significant difference between ROC curves was computed using the roc.test() command.

RESULTS

RRBS profiling of *ZNF154* methylation density in healthy and cancerous tissues

We used the *ZNF154* locus as our model cancer methylation biomarker given its propensity to be specifically and recurrently hypermethylated in numerous cancer types as compared to healthy control tissues, including white blood cells (WBCs), the predominant source of cfDNA in liquid biopsy specimens (32). As shown in Figure 1A, data derived from analysis of publicly-available Illumina Infinium HumanMethylation450 BeadChip data [including Widschwendter *et. al.* (8), Lehne *et. al*. (22) and The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/)] demonstrated cancer-specific hypermethylation throughout the entire *ZNF154* CpG island (CGI), and most differentially near the 5' end of the gene. Therefore, we focused our study on a 175-bp locus covering this region of the CGI (highlighted in yellow, Figure 1A). We next performed an initial analysis of RRBS data (8) generated from EOCs, healthy ovarian tissues, and WBCs. We defined methylation density as the ratio of methylated CpG dinucleotides (meCpGs) to total CpG sites within a given DNA molecule (i.e., epiallele) from a defined genomic locus. For comparison, the mean locus methylation was calculated as the number of meCpGs over all CpGs covered by the epialleles derived from the locus in question, as previously defined (24). According to the methylation density profiles of the three samples types, (Figure 1B, C, and D; depicting the relative prevalence of epiallelic fractions as a function of locus methylation density) there was a higher proportion of heavily-methylated epialleles in EOC samples when compared to either control or WBC samples.

Development and performance of a methylation density binary classifier in tissue

We developed a basic framework for a methylation density binary classifier (MDBC; see Figure 2A for schematic), aimed at identifying ideal parameters for classifying samples according to their respective methylation density profiles. In short, the MDBC incorporates an iterative process to identify cutoff values for methylation density and epiallelic fractions that maximize discrimination (true positive rate [TPR] minus

false positive rate [FPR]) between two defined cohorts. This is achieved by the incremental application of a methylation density cutoff to define sample epiallelic fractions followed by the iterative calculation of the TPR-FPR values for the cohorts as a function of the epiallelic fraction cutoff. In this approach a sample is called positive if its epiallelic fraction is above the candidate cutoff values. The resulting data can then be used to identify parameters to achieve the maximum TPR – FPR value based on a given methylation density dataset.

We first investigated the potential of the MDBC to identify optimal cutoff values using the RRBS data generated from the three aforementioned sample types. For discriminating EOCs from healthy ovarian tissue or WBC controls, the MDBC identified optimal methylation density cutoff values of 40% or 85%, and optimal epiallelic fraction cutoff values of 7.9% or 1.1%, respectively. Figure 2 B-C contrasts the performance of the MDBC and classification by mean locus methylation with respect to receiver operating characteristic (ROC) analysis and sample classification. We found that the MDBC provided a marginal improvement over mean methylation with respect to AUC values but no improvement for discriminating EOC from healthy ovarian tissues (at 100% specificity). In contrast, when comparing *ZNF154* methylation in EOCs to WBCs, our use of MDBC provided considerably better discriminatory power, as evidenced by a substantially lower p-value of 1.41e-5, compared to 7.57e-3 when measured by mean locus methylation. Here, the measurement of respective (densely-methylated) epiallelic fractions provided improved discriminatory potential to classify samples and improve the clinical sensitivity (i.e., TPR) by a factor of roughly 30% (66% vs. 92%), while also maintaining absolute clinical specificity (i.e., 100% true negative rate).

MDBC performance on simulated dilute admixtures

We next explored how the MDBC might be expected to perform in dilute admixture solutions, such as cfDNA. Toward this end, we created simulated cfDNA solutions using *in silico* admixtures from RRBS reads of EOCs (n=12) and WBCs (n=22). Positive cfDNA samples were simulated by combining 10,000 randomly sampled RRBS reads from each of the EOC samples after being randomly-paired and mixed with a WBC background sample at various EOC:WBC admixture ratios ranging from 100% down to 0.01%. These were then compared with the RRBS reads exclusively from the WBCs. The methylation density profiles from the

WBCs and the simulated positive cfDNA samples were then used, as above, to calculate the discriminatory power (TPR – FPR) of the MDBC over the entire range of methylation density and epiallelic fraction cutoff values. Data from this evaluation were then used to identify those methylation density cutoff values achieving the highest AUC value based on the MDBC at each admixture ratio. Lastly, the AUC values achieved by implementation of the MDBC to classify the samples were compared to those achieved using mean locus methylation.

Figure 3A shows the results of the AUC performance of MDBC versus average methylation classification over the entire range of simulated EOC:WBC admixture ratios. The results of this analysis indicate that MDBC classification outperformed average methylation at all admixture ratios, particularly as the EOC:WBC ratio approached the 0.01-1% range commonly reported for ctDNA in cfDNA from cancer patients (33). To further evaluate the predicted MDBC performance in liquid biopsies, we generated a second simulated cohort of 50 cancer-positive and 50 cancer-free cfDNA samples, to observe the variance in MDBC performance in simulated cfDNA solutions of 1%, 0.1% and 0.01% EOC:WBC admixture ratios. The results of these simulations, shown in Figure 3B and Supplemental Figure S1, indicate that while the probability of improving sample classification relative to mean locus methylation steadily increases as the methylation density cutoff increases from 20 to approximately 80% at all admixture ratios, the anticipated performance improvement is less significant as the admixture ratio approaches 0.01%, as the ctDNA fraction becomes only stochastically present.

Implementation of the MDBC with DREAMing in liquid biopsies

We next aimed to validate the utility of the MDBC by assessing its performance for classifying samples based on *ZNF154* methylation profiles obtained directly from liquid biopsy samples. To determine these profiles, we first designed and optimized a DREAMing-based assay targeting 14 CpGs located in the 175-bp segment of the *ZNF154* CGI (Figure 1A). We validated the performance of this assay by performing quasi-digitization of discrete copy numbers of synthetic equivalents of bisulfite treated (BST) methylated epialleles in a large excess of BST unmethylated genomic DNA and observed the resulting DREAMing high resolution melt (HRM) profiles, as previously described (21). The results of the validation demonstrated the anticipated Poissonian distribution of wells exhibiting a secondary negative derivative melt profile peak

(melt temperature, T_m) between 83.6°C and 84.0°C, indicative of single copy sensitivity for fully methylated *ZNF154*, while the remainder of the wells exhibited a single melt peak between 80.0°C to 80.4°C temperature, ostensibly derived exclusively from unmethylated genomic DNA (Supplementary Figure S2).

We next employed the *ZNF154* DREAMing assay to assess *ZNF154* methylation density profiles in a cohort of liquid biopsies from 71 patients (26 women with late-stage ovarian cancer and 45 cancer-free women; Supplementary Table S1). Each sample was first divided into a sufficient number of wells to achieve quasi-digitization based upon an anticipated epiallelic fraction of 0.01 – 0.1%. Following PCR amplification and HRM, the melt temperature of each detected epiallele was determined and enumerated for each respective sample. We performed next-gen amplicon bisulfite sequencing of 77 wells encompassing 131 epialleles, identified as positive by DREAMing analysis, in order to validate and calibrate the linearity of the relationship between the amplicon melt temperatures and epiallelic methylation density (Supplementary Figure S3). Overall, results of this analysis demonstrated that melt temperatures determined by DREAMing correlated linearly with the methylation density of the corresponding epialleles. Based on this calibration, methylation density profiles were then created for each liquid biopsy specimen (Supplementary Table 1). As shown in Figure 4A, preliminary meta-analysis of the methylation density profiles obtained via DREAM analysis of EOC-positive and cancer-free women revealed that while both cohorts exhibited a large fraction of *ZNF154* epialleles with little to no methylation, only cfDNA from cancer-positive patients exhibited the presence of a significant proportion of densely-methylated *ZNF154* epialleles.

We next employed the MDBC to identify methylation density and epiallelic fraction cutoffs which would optimize the diagnostic performance for classifying EOC-positive women based solely on their respective liquid biopsy *ZNF154* methylation density profiles. The MDBC analysis, shown in Figure 4B, identified 45% and 0.14% as the optimal cutoff values for methylation density and epiallelic fraction, respectively. Figure 4C shows the classification performance as a function of methylation density threshold (using corresponding optimal epiallelic cutoffs) in comparison to classification by mean methylation alone. Data from this analysis demonstrate that overall performance increases as the methylation density threshold is increased to 20% and remains largely flat until reaching 90%, at which point it declines, indicating that consideration of heterogeneously methylated epialleles can significantly improve

classification performance over assessment of only heavily-methylated epialleles, as typically determined by MSP-based assays. In contrast, consideration of epialleles at methylation densities below 20% reduces overall performance by effectively reducing assay specificity. These points are further illustrated in Figures 4D, which shows that the MDBC achieves significantly better ROC performance than classification using either a 0% or 90% methylation density threshold. Figure 4E illustrates that implementation of the optimal thresholds is able to achieve a sensitivity and specificity of 73.1% and 95.6%, respectively, compared with 65.4% and 77.8% via discrimination by mean methylation, inferred from considering all epialleles, and 53.8% and 100% if only heavily methylated epialleles are considered.

Independent validation of MDBC threshold values

We validated the methylation density and epiallelic fraction thresholds identified by the MDBC from our initial cohort by applying them to the analysis of a second, independent cohort. For this validation cohort, we used archived plasma samples (n=36), comprising 3 separate blood draws obtained from 8 women over a period of 8 weeks (n=24), as well as plasma samples obtained from cancer-free women (n=12; Supplementary Table S2). As before, plasma-derived cfDNA from each sample was analyzed via *ZNF154* DREAMing assay to determine its corresponding methylation density profile. Initial analysis, shown in Figure 5A, indicated that these samples exhibited a similar overall methylation density profile as the first cohort. We then blindly classified these samples according to the 45% optimal methylation density and 0.14% epiallelic fraction thresholds established by MDBC analysis of the first sample cohort, the results of which are shown in Figure 5B and Supplementary Figure S3. Overall, implementation of these threshold values performed similarly well in the validation cohort, achieving a classification performance of 87.5% sensitivity and 91.7% specificity.

Independent evaluation of the DREAMing data from the validation cohort, shown in Supplementary Figure S4, demonstrated that the ideal methylation density cutoff values identified by the MDBC were highly similar for both the initial (20-50% density) and validation (20-35% density) cohorts. These results indicate that experimentally-determined methylation density thresholds are likely consistent and applicable when evaluating independent cohorts with similar patient characteristics.

Lastly, we compared the diagnostic performance of the MDBC optimized cutoffs with commonly-employed CA-125 diagnostic criteria (34) for blood-based detection and monitoring of ovarian cancer. In the second plasma cohort, CA-125 measurements were available for all but one of the blood draws. The MDBC cutoffs correctly classified 20 of 23 (87.0%) samples, significantly exceeding the performance of CA-125, with which only 11 of 23 (47.8%) samples were classified correctly using the standard cutoff of 35 U/mL (Figure 5D). More specifically, CA-125 misclassified all of the patients with non-serous ovarian cancer subtypes (n=9). By contrast, employment of the MDBC with *ZNF154* DREAMing data correctly classified 7 of the 9 non-serous samples. This finding suggests that optimizing the diagnostic performance of *ZNF154* methylation by way of the MDBC may represent a substantial improvement over CA-125, particularly for women with non-serous ovarian cancer.

DISCUSSION

In the present work, we report a method for leveraging disease-associated differences in epiallelic methylation density profiles to improve performance of methylation biomarkers for use in clinical diagnostic assays, particularly when evaluating challenging samples such as liquid biopsies. Our results demonstrate that assessing methylation density information of individual DNA fragments can be used to establish effective thresholds capable of overcoming the inherent biological noise associated with methylation from background sources such as healthy tissues or age-related epigenetic drift (35). Using the methylation biomarker *ZNF154* in the context of ovarian carcinoma as an example, we demonstrated that methylation density profiles can be inferred from sequencing data from tissues to predict performance of candidate methylation biomarkers. We employed an ultra-sensitive technique called DREAMing to determine the methylation density of individual DNA fragment epialleles in plasma at this locus and used the resulting information to validate the effectiveness of classifying samples based upon their respective *ZNF154* methylation density profiles.

While the MDBC can be used to predict and improve the performance of newly-identified methylated biomarkers, it also has the potential to prompt reevaluation of promising methylated loci that may have been overlooked or excluded due to perceived background noise resulting from heterogeneous methylation. As this noise depends not only on the locus in question but also on the cohort of samples

analyzed, the MDBC could be used to prescreen for methylated loci that would have the highest performance specific to the clinical context of interest. Moreover, the optimal methylation density cutoff could also be used to inform the type of assay to be designed for the identified methylated biomarker. For example, an identified locus with a high methylation density optimal cutoff would imply that an MSP-based assay targeting these epialleles would be sufficient for diagnostic purposes instead of developing a more in-depth sequencing approach.

We observed that the optimal thresholds identified by the MDBC can vary considerably between distinct datasets as seen in the notable difference in methylation density thresholds between the ovarian tissue and plasma sample sets. This likely reflects inherent differences in the composition of the tissue types and suggests that care should be taken when employing MDBC-identified thresholds in disparate sample types. Nonetheless, our results indicate that consideration of heterogeneously methylated epialleles can likely improve diagnostic performance over a wide range of threshold values, regardless of sample type or cohort.

Recent analyses of cancer methylomes have shown that while cancer-specific hypermethylation can be highly deterministic, methylation patterns between tumor subpopulations at these loci can also be highly polymorphic (13). Methylation assays, such as MSP, that rely on heavily methylated signatures may become ineffective at detecting early stage tumors in part due to their inability to discriminate subtle changes in methylation density that likely arise early during carcinogenesis. In theory, MDBC thresholds can potentially be adjusted to differentiate even marginal changes in methylation density to increase early stage sensitivity while balancing specificity. Future studies will be needed to establish the relationship between the extent of heterogeneous methylation at each tumor stage across genomic loci and the optimal MDBC cutoffs for each.

The MDBC approach is critically reliant upon sensitive and accurate assessment of methylation density at the individual epiallele level. This can pose technical or logistic barriers, particularly for sequencing-based approaches, which require significant time and cost to achieve adequate sensitivity and statistical power to determine accurate methylation density profiles in samples containing dilute tumor DNA, such as liquid biopsies. Alternatively, the DREAMing-based system we describe here offers several advantages for targeted profiling of methylation density. In particular, the short turnaround time for

DREAMing (results in several hours) and low cost (approximately $10.00 per sample) make it a practical option for profiling methylation density. Secondly, single molecule sensitivity is readily achievable by DREAMing, which is particularly important when working, as here, with limited (plasma) sample volumes. Furthermore, unlike bisulfite sequencing, DREAMing does not require analysis of sequencing results or patterns, drastically reducing the turnaround time for determining optimal thresholds for maximizing performance or applying established thresholds in a given clinical application.

There are a number of drawbacks worth considering when employing the DREAMing technique. Perhaps the most notable limitation of DREAMing is its reliance upon nonspecific DNA-binding dyes that all but preclude its amenability to multiplex analysis or large biomarker panels. Accurate enumeration of heterogeneously-methylated epialleles also the requires that each sample undergo quasi-digital dilution to enable Poissonian quantitation, which ultimately limits target quantification to three orders of magnitude or less when using standard microtiter plates. Nonetheless, for the vast majority of applications, it is the rarer, more densely-methylated epialleles that are of primary interest. Taking this into account, selective absolute quantification can be achieved by performing limiting dilution targeted to only rare epialleles over a specific methylation density cutoff. This thereby requires less partitioning and diluting of samples and increases overall throughput. More recently, we have successfully incorporated DREAMing into a highly-parallelized microfluidic array that drastically enhances the dynamic range of the assay, allowing simultaneous methylation density analysis of millions of molecules per sample, obviating the need to quantify sample DNA concentrations to achieve precise dilution (36). Lastly, it is worth noting that the exquisite sensitivity and resolution of DREAMing make it potentially susceptible to issues such as inefficient bisulfite conversion or low fidelity polymerases that can lead to amplicons that do not accurately reflect the sequences or methylation statuses of the template epialleles. Likewise, it is strongly recommended to use only established and validated protocols and reagents for sample processing and analysis.

In the present work, we employed *ZNF154* as a model methylation biomarker based upon previous reports by us and others showing it to be recurrently methylated in numerous cancer types. The findings in our present study confirm and extend these observations by demonstrating that methylation of *ZNF154* can reliably distinguish cfDNA from liquid biopsies of EOC-positive women from those of healthy, noncancerous controls. Furthermore, our results also indicate that, unlike CA-125, the broad cancer specificity of *ZNF154*

methylation can be leveraged for the detection of all EOC subtypes, including non-serous cases. On the other hand, while the broad cancer specificity of *ZNF154* makes it attractive as a potential pan-cancer biomarker, it would likely need to be paired with follow up analyses such as a panel of other, tissue-specific markers to identify the cancer tissue of origin (18,37).

Overall, the MDBC approach presented here is a relatively simple, but effective technique for overcoming issues related to the heterogeneous methylation patterns that arise from the infidelity and stochasticity of the cellular processes associated with the accumulation and maintenance of DNA methylation. Likewise, this approach is likely to be largely suitable not only for optimizing the performance of methylation biomarkers for clinical diagnostic applications, but also more generally for the study and evaluation of any biological phenomena associated with the dynamic accumulation or loss of DNA methylation.

DATA AVAILABILITY

A 'beta version' of the MDBC Python code and input files is available at the GitHub website: https://github.com/bmill3r/MDBC. This GitHub project is intended to demonstrate the steps to generate the analyses and figures presented in this study. Work is underway to extend the program for general use with other datasets and genomic loci. Interested parties may contact the authors for more details.

FUNDING

Funding provided by the Intramural program of the National Human Genome Research Institute; the Intramural Research Program of the NCI; the Johns Hopkins Kimmel Cancer Center-Allegheny Health Network; and the National Institutes of Health (R01CA155305, R21CA186809, U54CA151838 and U01CA214165).

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENT

We would like to thank Kristin Harper and Karoun Bagamian for helpful edits and useful feedback during the writing of this piece, Ryan Winters for supplying us with plasma samples from the Fox Chase Cancer Biorepository, Alice Young for guidance during sequencing library preparation and data acquisition, Dr. Martin Widschwendter for access to the RRBS datasets used in this study, Alejandro Stark for instruction on cfDNA extraction and processing, and Thor Nilsen for technical assistance with the NeoGeneStar cfDNA Purification Kit.

AUTHOR CONTRIBUTIONS

Conceptualization: BFM, TRP, LE; Experiments: BFM, TRP, HP, PA; Data Analysis and Visualization: BFM, GM, AO, CMA; Writing – Original Draft: BFM, TRP; Writing – Review and Editing: BFM, TRP, LE, Funding Acquisition: LE, TRP, TW, CMA; Supervision: LE, TRP; Samples: AO, CMA

REFERENCES

1. Sanchez-Vega, F., Gotea, V., Margolin, G. and Elnitski, L. (2015) Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype. *Epigenetics Chromatin*, **8**, 14.
2. Sanchez-Vega, F., Gotea, V., Petrykowska, H.M., Margolin, G., Krivak, T.C., DeLoia, J.A., Bell, D.W. and Elnitski, L. (2013) Recurrent patterns of DNA methylation in the ZNF154, CASP8, and VHL promoters across a wide spectrum of human solid epithelial tumors and cancer cell lines. *Epigenetics*, **8**, 1355-1372.
3. Herman, J.G. and Baylin, S.B. (2003) Gene Silencing in Cancer in Association with Promoter Hypermethylation. *New England Journal of Medicine*, **349**, 2042-2054.
4. Crowley, E., Di Nicolantonio, F., Loupakis, F. and Bardelli, A. (2013) Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*, **10**, 472-484.
5. Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477.

6. Zhou, W., Laird, P.W. and Shen, H. (2017) Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic acids research*, **45**, e22-e22.

7. Mikeska, T. and Craig, J.M. (2014) DNA methylation biomarkers: cancer and beyond. *Genes (Basel)*, **5**, 821-864.

8. Widschwendter, M., Zikan, M., Wahl, B., Lempiainen, H., Paprotka, T., Evans, I., Jones, A., Ghazali, S., Reisel, D., Eichner, J. *et al.* (2017) The potential of circulating tumor DNA methylation analysis for the early detection and management of ovarian cancer. *Genome Med*, **9**, 116.

9. Campan, M., Moffitt, M., Houshdaran, S., Shen, H., Widschwendter, M., Daxenbichler, G., Long, T., Marth, C., Laird-Offringa, I.A., Press, M.F. *et al.* (2011) Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. *PLoS One*, **6**, e28141.

10. Fiala, C., Kulasingam, V. and Diamandis, E.P. (2018) Circulating Tumor DNA for Early Cancer Detection. *The Journal of Applied Laboratory Medicine*, **3**, 300-313.

11. Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D. *et al.* (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*, **43**, 768-775.

12. Timp, W. and Feinberg, A.P. (2013) Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nature reviews. Cancer*, **13**, 497-510.

13. Landan, G., Cohen, N.M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D.A., Goldfinger, N., Zundelevich, A. *et al.* (2012) Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet*, **44**, 1207-1214.

14. Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N. *et al.* (2017) A DNA methylation map of human cancer at single base-pair resolution. *Oncogene*, **36**, 5648-5657.

15. Field, A.E., Robertson, N.A., Wang, T., Havas, A., Ideker, T. and Adams, P.D. (2018) DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Molecular Cell*, **71**, 882-895.

16. Devonshire, A.S., Whale, A.S., Gutteridge, A., Jones, G., Cowen, S., Foy, C.A. and Huggett, J.F. (2014) Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal Bioanal Chem*, **406**, 6499-6512.

17. Shen, S.Y., Singhania, R., Fehringer, G., Chakravarthy, A., Roehrl, M.H.A., Chadwick, D., Zuzarte, P.C., Borgida, A., Wang, T.T., Li, T. *et al.* (2018) Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature*, **563**, 579-583.

18. Lehmann-Werman, R., Neiman, D., Zemmour, H., Moss, J., Magenheim, J., Vaknin-Dembinsky, A., Rubertsson, S., Nellgard, B., Blennow, K., Zetterberg, H. *et al.* (2016) Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A*, **113**, E1826-1834.

19. Guo, S., Diep, D., Plongthongkum, N., Fung, H.L., Zhang, K. and Zhang, K. (2017) Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*, **49**, 635-642.

20. Margolin, G., Petrykowska, H.M., Jameel, N., Bell, D.W., Young, A.C. and Elnitski, L. (2016) Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development. *J Mol Diagn*, **18**, 283-298.

21. Pisanic, T.R., 2nd, Athamanolap, P., Poh, W., Chen, C., Hulbert, A., Brock, M.V., Herman, J.G. and Wang, T.H. (2015) DREAMing: a simple and ultrasensitive method for assessing intratumor epigenetic heterogeneity directly from liquid biopsies. *Nucleic Acids Res*, **43**, e154.

22. Lehne, B., Drong, A.W., Loh, M., Zhang, W., Scott, W.R., Tan, S.T., Afzal, U., Scott, J., Jarvelin, M.R., Elliott, P. *et al.* (2015) A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*, **16**, 37.

23. Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571-1572.

24. Schultz, M.D., Schmitz, R.J. and Ecker, J.R. (2012) 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet*, **28**, 583-585.

25. Mason, S.J. and Graham, N.E. (2002) Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, **128**, 2145-2166.

26. Keeley, B., Stark, A., Pisanic, T.R., 2nd, Kwak, R., Zhang, Y., Wrangle, J., Baylin, S., Herman, J., Ahuja, N., Brock, M.V. *et al.* (2013) Extraction and processing of circulating DNA from large sample volumes using methylation on beads for the detection of rare epigenetic events. *Clin Chim Acta*, **425**, 169-175.

27. Candiloro, I.L., Mikeska, T., Hokland, P. and Dobrovic, A. (2008) Rapid analysis of heterogeneously methylated DNA using digital methylation-sensitive high resolution melting: application to the CDKN2B (p15) gene. *Epigenetics Chromatin*, **1**, 7.

28. Snell, C., Krypuy, M., Wong, E.M., kConFab, i., Loughrey, M.B. and Dobrovic, A. (2008) BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast Cancer Res*, **10**, R12.

29. R Core Team. (2017) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. *et al.* (2011) Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, **12**, 2825−2830.

31. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. and Muller, M. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, **12**, 77.

32. Sun, K., Jiang, P., Chan, K.C., Wong, J., Cheng, Y.K., Liang, R.H., Chan, W.K., Ma, E.S., Chan, S.L., Cheng, S.H. *et al.* (2015) Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A*, **112**, E5503-5512.

33. Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E. *et al.* (2017) Direct detection of early-stage cancers using circulating tumor DNA. *Science Translational Medicine*, **9**, eaan2415.

34. Bottoni, P. and Scatena, R. (2015) In Scatena, R. (ed.), *Advances in Cancer Biomarkers: From biochemistry to clinic for a critical revision*. Springer Netherlands, Dordrecht, pp. 229-244.

35. Issa, J.-P. (2014) Aging and epigenetic drift: a vicious cycle. *The Journal of clinical investigation*, **124**, 24-29.

36. O'Keefe, C.M., Pisanic, T.R., 2nd, Zec, H., Overman, M.J., Herman, J.G. and Wang, T.H. (2018) Facile profiling of molecular heterogeneity by microfluidic digital melt. *Sci Adv*, **4**, eaat6459.

37. Moran, S., Martinez-Cardús, A., Boussios, S. and Esteller, M. (2017) Precision medicine based on epigenomics: the paradigm of carcinoma of unknown primary. *Nature Reviews Clinical Oncology*, **14**, 682.

TABLE AND FIGURES LEGENDS

Figure 1. Methylation density at the *ZNF154* genomic locus. **A**) The *ZNF154* gene is encoded on the reverse strand of Chromosome 19 and contains a 328-bp CpG island (CGI) that extends from the 5'-UTR through into the *ZNF154* gene body itself. Schematic showing average methylation at multiple CpG sites determined from Illumina Infinium HumanMethylation450 array data for tumor (*red lines*) and control (*blue lines*) tissues. The target locus assessed in the DREAMing assay is highlighted (*yellow*). **B-D**) Heatmaps showing the relative fractions and corresponding methylation density profiles derived from RRBS reads of the *ZNF154* target locus for ovarian carcinomas (n=12), healthy ovarian tissues (n=10), and WBCs (n=22). Abbreviations: ovals below the CpG island represent CpG positions; EOCs = epithelial ovarian carcinomas (n=221); WBCs = white blood cells; * indicates data taken from Widschwendter *et. al*; ** indicates data from Lehne *et. al*; remaining 4 letter acronyms correspond to TCGA tissue codes: BLCA = bladder carcinoma (tumors = 412, controls = 21); COAD = colon adenocarcinoma (tumors = 295, controls = 38); HNSC = head-neck squamous cell carcinoma (tumors = 528, controls = 50); LIHC = liver hepatocellular carcinoma (tumors = 377, controls = 50); LUSC = lung squamous cell carcinoma (tumors = 370, controls = 42); PRAD = prostate adenocarcinoma (tumors = 498, controls = 50); STAD = stomach adenocarcinoma (tumors = 395, controls = 2); UCEC = uterine corpus endometrial carcinoma (tumors = 431, controls = 46).

Figure 2. MDBC classification of ovarian carcinomas using RRBS read data. **A**) Schematic illustrating the MDBC procedure for determining optimal methylation density and epiallelic fraction cutoffs**. B-C**) MDBC heatmaps indicating the true and false positive rate differences for each combination of methylation density and epiallelic fraction cutoffs for identification of EOCs versus healthy ovarian tissue or WBC controls. Aqua rectangles highlight the optimal methylation density and epiallelic fraction cutoffs. The maximum true and false positive rate difference is indicated with a black arrow. The performance of the optimal methylation density cutoff is shown as a ROC curve below the MDBC heatmap and the performance of the optimal epiallelic fraction cutoff at the optimal methylation density is shown as a boxplot. Abbreviations: TPR = true positive rate; FPR = false positive rate; MD = methylation density; EF = epiallelic fraction; AUC = area under the curve; MDBC = methylation density binary classifier; EOC = epithelial ovarian carcinomas (n=12); Healthy OV = healthy ovarian tissue control samples (n=10); WBC = white blood cell control samples (n=22); *, **, and *** indicates $p < 0.05$, $p < 0.01$, and $p < 0.001$, two-sided Wilcoxon rank-sum test.

Figure 3. Simulated performance of the MDBC using varying admixture ratios of ovarian carcinoma (EOC) to WBC RRBS reads. **A**) The performance of the methylation density binary classifier (MDBC, *red*) and mean locus methylation classifier (*blue*) at increasing dilutions of EOC RRBS reads in a background of WBC RRBS reads acquired from Widschwendter *et al*. **B**) Plots showing the probability of achieving a higher AUC than the mean locus methylation classifier using the MDBC for each methylation density cutoff at various admixture ratios. Lower panels show the TPR, 1-FPR, and AUC achieved by using the MDBC at each methylation density cutoff. Solid lines indicate the mean value and shaded regions indicate the 95% confidence interval for 50 iterations of the simulation. Abbreviations: AUC = area under the curve; MDBC = methylation density binary classifier; TPR = true positive rate; FPR = false positive rate.

Figure 4. Performance of the MDBC in EOC patient and control plasma samples. **A**) The pooled epiallelic fractions of cfDNA methylated epialleles with varying methylation densities in EOC (*red*, n=26) and healthy (*blue*, n=45) patient plasma samples. Purple shaded regions indicate overlap between the two plasma sets. **B**) MDBC heatmap indicating the true and false positive rate differences for each combination of methylation density and epiallelic fraction cutoffs for identification of EOC versus healthy control plasma

samples. **C**) Performance of the MDBC at each methylation density cutoff for the EOC and healthy control plasma samples. Dotted line shows the AUC of the mean methylation cutoff. **D**) ROC curves showing the classification performance of using the optimal methylation density cutoff determined by the MDBC (*red*), 90% methylation density cutoff (*yellow*), or mean methylation cutoff (*blue*) to identify the EOC and healthy control plasma samples. **E**) Boxplots showing the performance of the optimal epiallelic fraction cutoffs for either the optimal 45% methylation density cutoff determined by the MDBC (*red*), 90% methylation density cutoff (*yellow*), or mean methylation cutoff (*blue*) to classify the EOC patient (*red*, n=26) or healthy control (*blue*, n=45) plasma samples. Dotted lines indicate the optimal epiallelic fraction cutoff for each of the methylation density cutoffs. Mean methylation was inferred from the fraction of all methylated epialleles. Abbreviations: EOC = epithelial ovarian carcinoma; TPR = true positive rate; FPR = false positive rate; AUC = area under the curve; MDBC = methylation density binary classifier; MD = methylation density cutoff; * and *** indicates $p < 0.05$ and $p < 0.001$, two-sided Wilcoxon rank-sum test.

Figure 5. Validation of MDBC cutoff values and corresponding performance in comparison to CA-125 assessment for identifying EOC from patient plasma. **A**) The pooled epiallelic fractions of cfDNA methylated epialleles with varying methylation densities in the second EOC (*red*, n=24) and healthy (*blue*, n=12) patient plasma sample cohort. Purple shaded regions indicate overlap between the two plasma sets. **B**) Receiver operating characteristic curve for the optimal 45% methylation density cutoff on the second plasma cohort. **C**) Boxplots showing the performance of the optimal epiallelic fraction (*dotted line*; 0.14%) and methylation density (45%) cutoffs to classify the second EOC patient (*red*, n=24) and healthy control (*blue*, n=12) plasma sample cohort. **D**) Fractions of non-serous or serous EOC subtype patient plasma samples from the second cohort above the CA-125 or MDBC cutoffs. Abbreviations: EOC = epithelial ovarian carcinoma; MDBC = methylation density binary classifier; AUC = area under the curve; *** indicates $p < 0.05$, two-sided Wilcoxon rank-sum test.

Fig 1

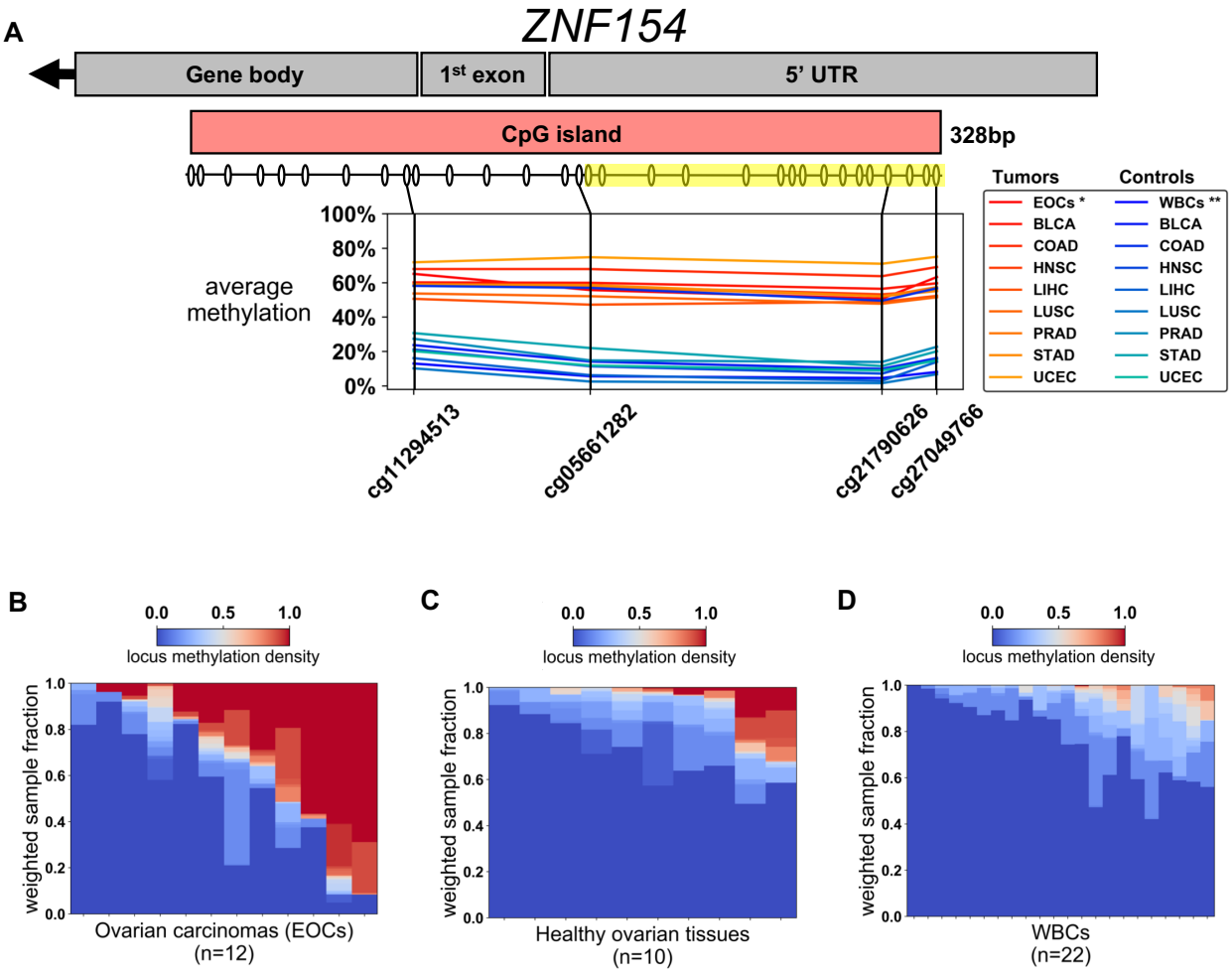
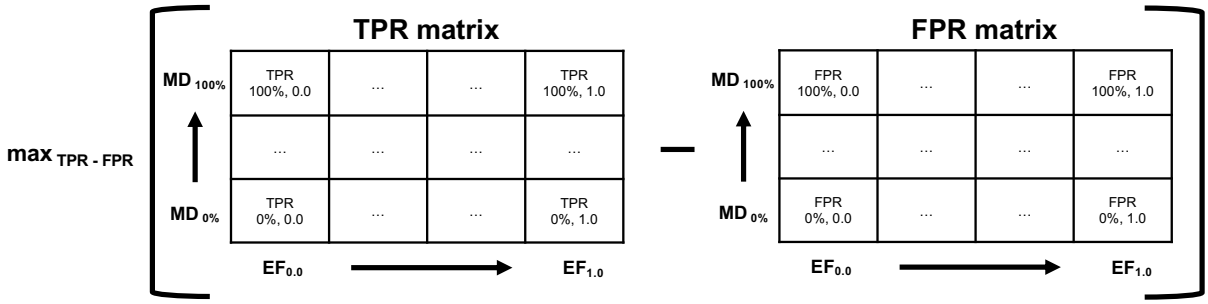
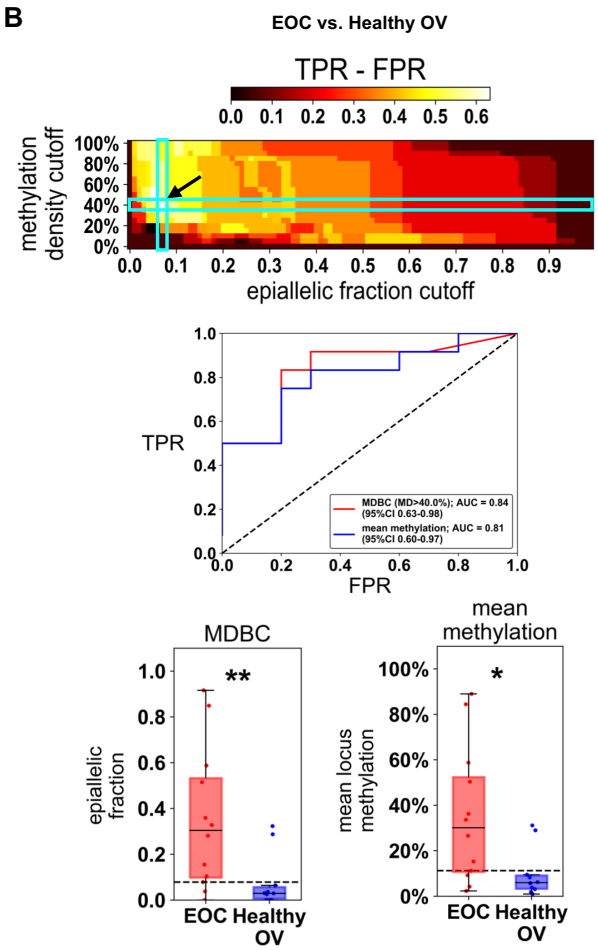

A
ZNF154
Gene body
1st exon
5' UTR
CpG island
328bp
average methylation
100%
80%
60%
40%
20%
0%
cg11294513
cg05661282
cg21790626
cg27049766
Tumors
Controls
EOCs *
BLCA
COAD
HNSC
LIHC
LUSC
PRAD
STAD
UCEC
WBCs **
BLCA
COAD
HNSC
LIHC
LUSC
PRAD
STAD
UCEC
B
C
D
0.0
0.5
1.0
locus methylation density
weighted sample fraction
Ovarian carcinomas (EOCs)
(n=12)
Healthy ovarian tissues
(n=10)
WBCs
(n=22)

Fig 2

A


TPR matrix
FPR matrix
max TPR - FPR
MD 100%
MD 0%
TPR 100%, 0.0
TPR 100%, 1.0
TPR 0%, 0.0
TPR 0%, 1.0
FPR 100%, 0.0
FPR 100%, 1.0
FPR 0%, 0.0
FPR 0%, 1.0
EF0.0
EF1.0

B


EOC vs. Healthy OV
TPR - FPR
methylation density cutoff
epiallelic fraction cutoff
TPR
FPR
MDBC (MD>40.0%); AUC = 0.84 (95%CI 0.63-0.98)
mean methylation; AUC = 0.81 (95%CI 0.60-0.97)
MDBC
mean methylation
epiallelic fraction
mean locus methylation
EOC
Healthy OV

C

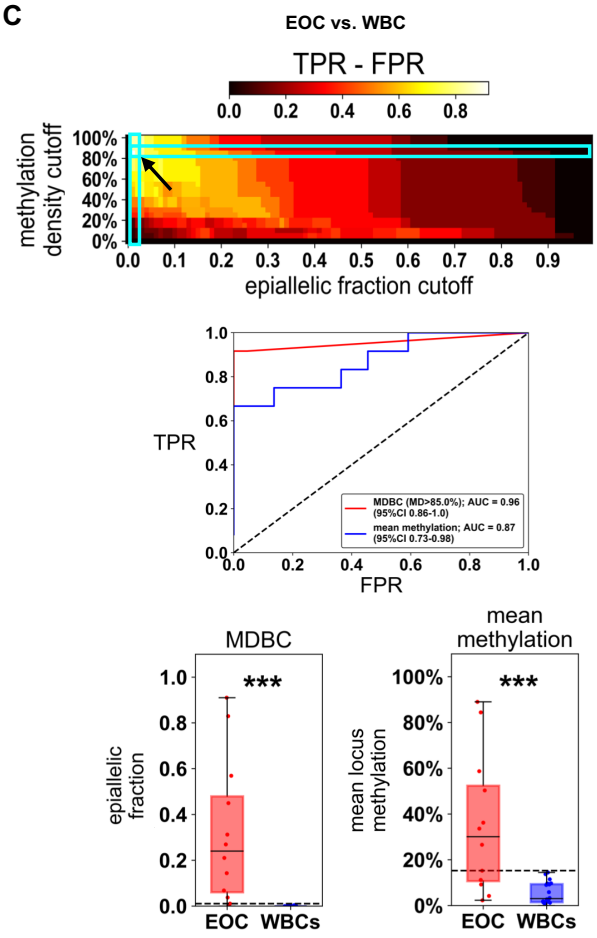

EOC vs. WBC
TPR - FPR
methylation density cutoff
epiallelic fraction cutoff
TPR
FPR
MDBC (MD>85.0%); AUC = 0.96 (95%CI 0.86-1.0)
mean methylation; AUC = 0.87 (95%CI 0.73-0.98)
MDBC
mean methylation
epiallelic fraction
mean locus methylation
EOC
WBCs

Fig 3

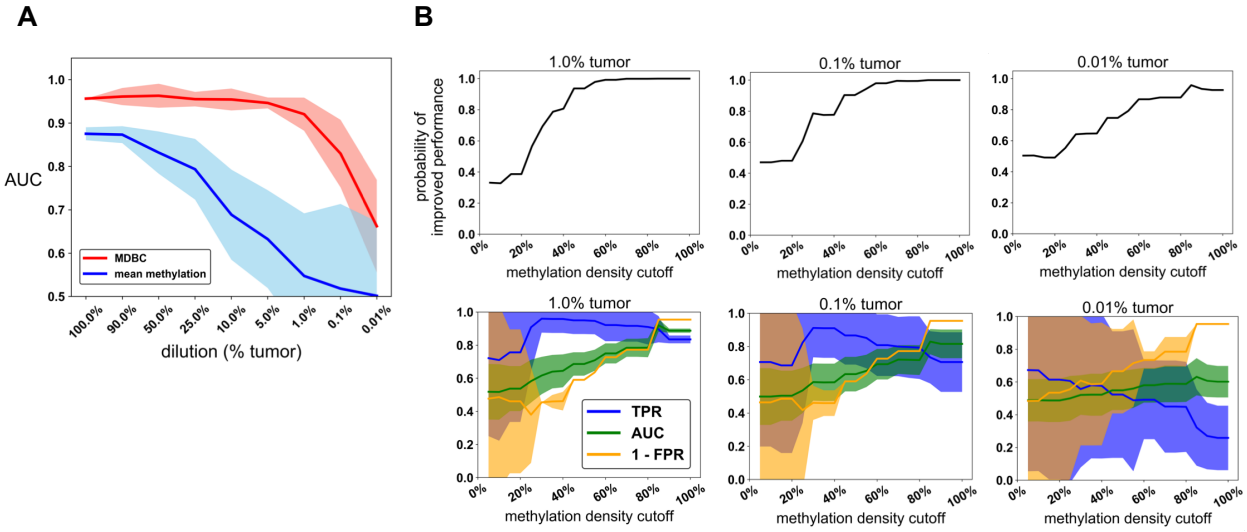

A
B
AUC
MDBC
mean methylation
dilution (% tumor)
100.0%
90.0%
50.0%
25.0%
10.0%
5.0%
1.0%
0.1%
0.01%
probability of improved performance
1.0% tumor
0.1% tumor
0.01% tumor
methylation density cutoff
TPR
AUC
1 - FPR

Fig 4

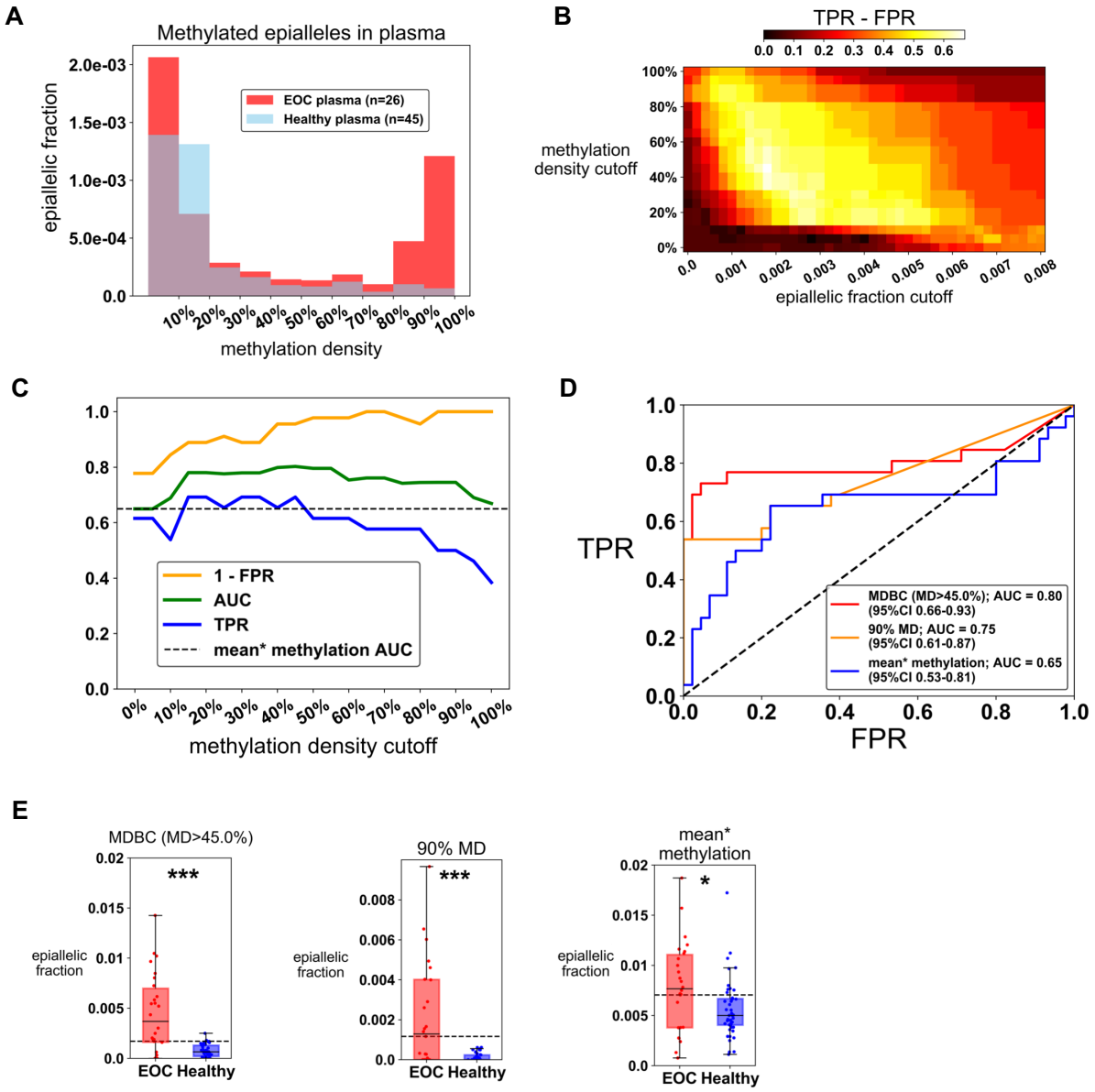
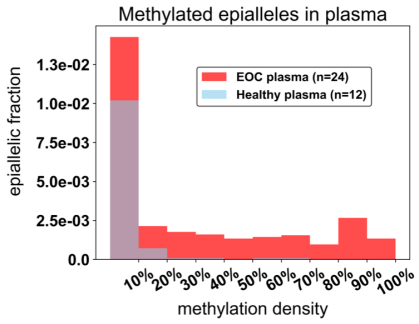
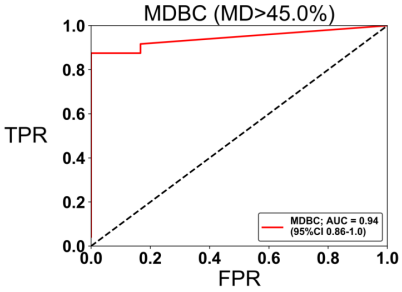

A
Methylated epialleles in plasma
EOC plasma (n=26)
Healthy plasma (n=45)
epiallelic fraction
2.0e-03
1.5e-03
1.0e-03
5.0e-04
0.0
10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
methylation density
B
TPR - FPR
0.0 0.1 0.2 0.3 0.4 0.5 0.6
methylation density cutoff
100% 80% 60% 40% 20% 0%
0.0 0.001 0.002 0.003 0.004 0.005 0.006 0.007 0.008
epiallelic fraction cutoff
C
1 - FPR
AUC
TPR
mean* methylation AUC
0.0 0.2 0.4 0.6 0.8 1.0
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
methylation density cutoff
D
TPR
FPR
0.0 0.2 0.4 0.6 0.8 1.0
MDBC (MD>45.0%); AUC = 0.80 (95%CI 0.66-0.93)
90% MD; AUC = 0.75 (95%CI 0.61-0.87)
mean* methylation; AUC = 0.65 (95%CI 0.53-0.81)
E
MDBC (MD>45.0%)
90% MD
mean* methylation
epiallelic fraction
***
*
0.0 0.005 0.01 0.015 0.02
0.0 0.002 0.004 0.006 0.008
EOC Healthy

Fig 5

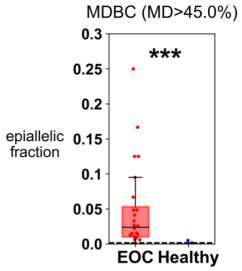
A



B



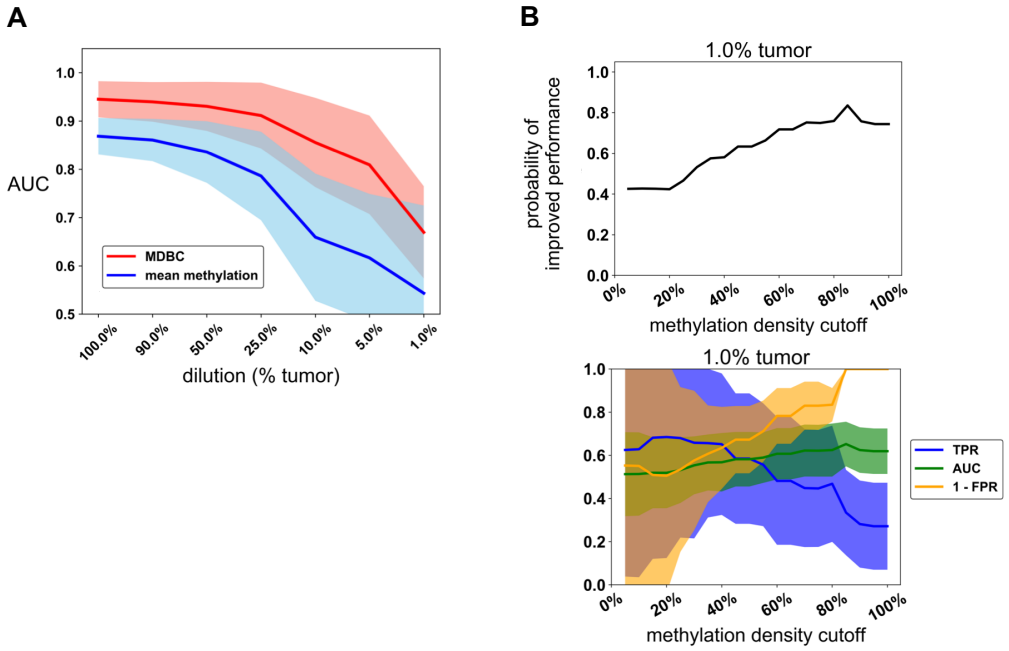
C



D

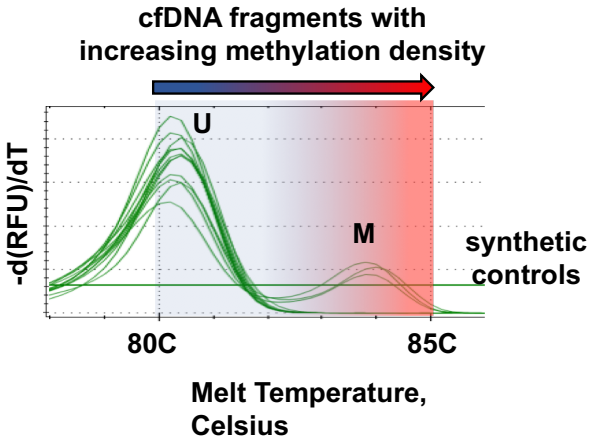
	Non-serous (n=9)	Serous (n=14)	Total (n=23)
CA125 (35U/mL cutoff)	0%	78.6%	47.8%
MDBC (MD>45% cutoff)	**77.8%**	**92.9%**	**87.0%**

Fig S1



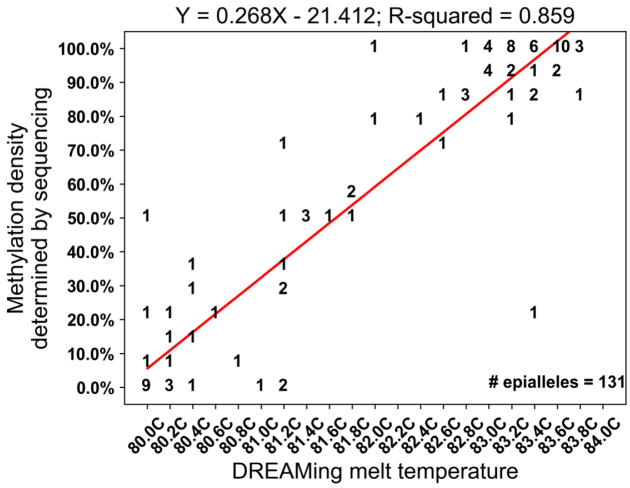
Supplemental Figure S1. Simulated performance of the MDBC using varying admixture ratios of ovarian carcinoma (EOC) to WBC RRBS reads sampled at 100 total reads per simulated sample. **A**) The performance of the methylation density binary classifier (MDBC, *red*) and mean locus methylation classifier (*blue*) at increasing dilutions of EOC RRBS reads in a background of WBC RRBS reads acquired from Widschwendter *et al*. **B**) Plots showing the probability of achieving a higher AUC than the mean locus methylation classifier using the MDBC for each methylation density cutoff at various admixture ratios. Lower panels show the TPR, 1-FPR, and AUC achieved by using the MDBC at each methylation density cutoff. EOCs (n=12) were randomly paired with a WBC (n=22) sample and RRBS reads were sampled from an EOC-WBC pair to generate a simulated spike-in sample. Simulated samples (n=12) were compared to the original WBC samples (n=22). 50 iterations of the simulation were performed. Solid lines indicate mean values and shaded regions indicate 95% confidence intervals. Abbreviations: MDBC = methylation density binary classifier; AUC = area under the curve; TPR = true positive rate; FPR = false positive rate. Abbreviations: AUC = area under the curve; MDBC = methylation density binary classifier; TPR = true positive rate; FPR = false positive rate.

Fig S2



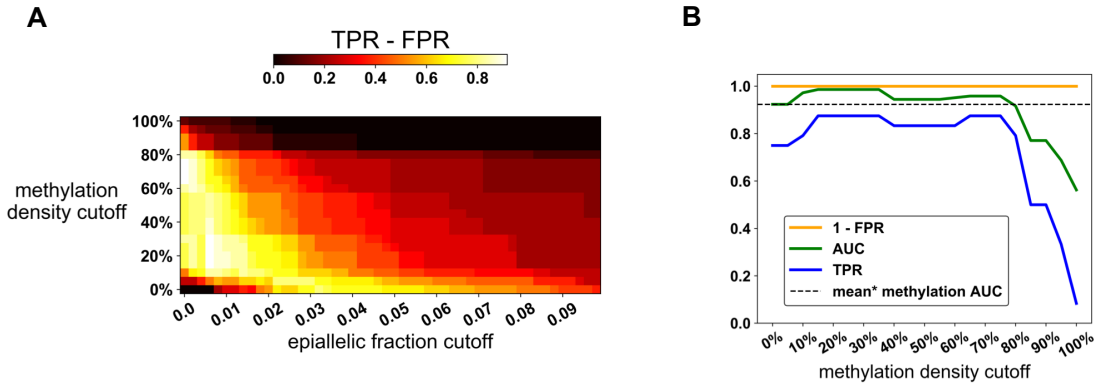
Supplemental Figure S2. Melting traces from DREAMing assay for wells containing a mixture of synthetic fully methylated (*M*) and unmethylated male sperm genomic DNA (*U*) fragments originating from the *ZNF154* genomic region of interest. Abbreviations: -d(RFU)/dT = negative derivative of the change in relative fluorescent units.

Fig S3



Supplemental Figure S3. *ZNF154* locus methylation density vs. DREAMing melt temperature in cfDNA. DREAMing melt peak temperatures and corresponding methylation density measurements identified via bisulfite sequencing for 131 post-DREAMing epiallele amplicons. The numbers on the plot represent the number of times a detected melt peak had the corresponding melt temperature and produced an amplicon with the corresponding methylation density. The red line denotes the best fit regression line. The linear regression model and R-squared value is shown above the plot.

Fig S4



Supplemental Figure S4. MDBC analysis of cfDNA DREAMing data from the validation cohort. **(A)** MDBC heatmap indicating the true and false positive rate differences for each combination of methylation density and epiallelic fraction cutoffs for identification of EOC patient (n=24) versus healthy control (n=12) plasma samples of the second cohort. **(B)** Performance of the MDBC at each methylation density cutoff for the plasma samples of the second cohort. * mean methylation determined from methylation density cutoff of 0%.

					DREAMing Profile (peak counts for each % methylation density)															
Sample	Age	Volume plasma (mLs)	ng/mL	genomic equivalents loaded	0%	7%	14%	21%	28%	35%	42%	49%	56%	63%	70%	77%	84%	91%	98%	100%
Ovarian 1	68	1.25	12.6	3528	5	9	11	1	0	1	1	1	2	0	0	0	0	1	0	0
Ovarian 2	77	1.25	18.7	4812	1	3	9	5	2	1	0	1	0	0	1	0	1	1	1	2
Ovarian 3	72	1.25	30.0	7452	2	3	1	0	1	3	0	0	0	0	0	1	5	8	1	7
Ovarian 4	27	1.25	38.7	8748	8	4	8	1	0	0	0	0	0	0	0	0	0	0	0	0
Ovarian 5	61	1.25	7.2	1992	9	3	2	0	1	0	0	0	0	0	0	0	2	0	3	3
Ovarian 6	59	1.25	18.3	4380	6	1	2	0	0	0	0	0	0	2	3	0	1	0	0	0
Ovarian 7	75	1.25	15.6	3924	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Ovarian 8	55	1.25	33.2	6000	1	2	2	0	0	1	0	0	1	1	0	0	5	4	11	3
Ovarian 9	40	1.00	62.3	12888	15	23	7	1	0	1	1	1	0	1	0	1	2	1	0	0
Ovarian 10	61	1.00	19.9	3624	3	1	4	1	0	0	0	0	0	0	0	0	0	0	0	0
Ovarian 11	51	1.00	24.1	4968	3	2	1	0	0	0	0	0	0	0	0	0	4	1	0	2
Ovarian 12	60	1.00	20.9	3216	4	9	0	1	1	1	0	1	4	0	0	1	2	0	0	0
Ovarian 13	61	1.00	38.6	9348	1	19	14	4	1	2	0	0	1	0	1	0	0	0	0	0
Ovarian 14	69	1.20	55.6	6000	18	2	0	0	0	0	0	0	0	0	0	0	0	0	7	17
Ovarian 15	63	1.00	18.3	3696	8	3	0	0	0	2	0	0	0	0	2	0	2	1	0	0
Ovarian 16	52	1.00	15.0	3120	8	5	1	2	0	3	1	1	1	0	2	0	0	1	0	0
Ovarian 17	54	1.00	6.0	1284	5	1	1	2	0	0	0	0	0	1	0	0	2	1	1	0
Ovarian 18	49	1.00	10.8	2244	3	0	2	0	0	3	0	0	0	0	0	0	2	9	10	3
Ovarian 19	53	1.00	10.5	2292	7	2	2	0	2	2	0	0	1	0	0	1	3	7	2	2
Ovarian 20	72	1.00	15.1	3036	4	4	5	0	1	1	0	1	0	0	0	1	3	2	3	3
Ovarian 21	66	1.00	19.2	4128	4	2	3	1	1	1	0	1	0	0	2	2	9	1	2	3
Ovarian 22	52	1.00	8.6	1920	3	2	1	1	0	0	0	2	0	0	0	1	1	1	4	0
Ovarian 23	33	1.00	14.8	3036	3	1	0	0	0	1	0	1	0	0	0	1	7	6	0	3
Ovarian 24	59	1.00	8.4	1968	6	4	3	1	2	1	2	0	2	0	0	0	0	0	0	0
Ovarian 25	67	1.00	34.9	6000	4	1	2	0	0	1	1	1	4	4	2	3	5	4	0	1
Ovarian 26	57	1.50	14.3	4584	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0
Normal 1	47	1.60	17.8	6240	4	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0
Normal 2	47	1.50	32.4	8760	4	6	4	2	0	3	1	2	3	0	0	1	2	0	0	0
Normal 3	48	1.50	31.6	9840	5	3	9	4	2	3	1	0	0	0	1	0	1	0	0	1
Normal 4	51	1.60	68.6	17484	0	4	12	7	1	7	2	3	3	1	2	1	1	0	0	0
Normal 5	63	1.60	33.6	9492	4	5	4	1	3	3	4	1	2	0	0	0	1	0	0	0
Normal 6	45	1.60	21.6	7236	4	9	7	2	3	7	0	1	1	0	0	0	3	1	0	1
Normal 7	48	1.80	30.0	9264	2	6	9	3	0	0	0	0	0	0	0	0	0	0	0	0
Normal 8	52	1.80	30.4	9268	1	2	10	2	2	2	0	0	0	1	0	1	0	0	0	0
Normal 9	53	1.80	16.5	7620	0	10	5	3	1	3	1	0	0	0	0	1	0	0	0	0
Normal 10	61	1.70	47.2	14064	4	3	13	4	1	1	0	1	0	0	0	0	0	0	2	0
Normal 11	65	1.80	159.7	37416	1	0	18	5	3	1	0	0	1	0	0	1	1	0	0	0
Normal 12	54	1.80	14.7	5580	6	11	7	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 13	60	2.00	18.8	6000	0	17	7	0	0	0	0	1	0	2	0	1	1	0	0	1
Normal 14	60	2.00	12.5	4836	7	10	6	0	0	0	0	0	0	2	0	0	0	0	0	1
Normal 15	59	1.80	31.4	12203	0	15	31	2	0	0	0	0	3	0	1	0	1	1	0	0
Normal 16	57	1.70	7.5	2580	7	8	1	0	0	0	1	0	0	0	1	0	0	0	1	0
Normal 17	54	1.70	13.8	4872	4	6	5	0	1	3	1	2	0	1	0	0	0	0	0	0
Normal 18	65	2.00	21.3	6000	10	8	4	0	0	1	1	1	1	0	1	1	0	0	0	1
Normal 19	57	2.20	10.5	4452	1	3	13	0	1	6	1	0	3	1	0	0	0	0	0	0
Normal 20	50	2.00	16.6	6000	3	9	9	2	1	0	0	0	0	0	0	0	0	0	0	0
Normal 21	55	1.90	15.5	5364	3	16	5	0	0	0	0	0	0	0	0	0	1	0	0	0
Normal 22	55	2.00	66.0	16000	0	3	48	7	2	1	0	0	1	1	3	2	1	2	1	0
Normal 23	58	2.00	10.8	3996	3	9	3	1	0	1	0	0	0	1	0	0	0	1	0	0
Normal 24	58	2.00	17.8	6132	6	3	10	0	0	1	0	1	0	0	0	0	0	0	0	1
Normal 25	64	2.00	502.7	18712	0	1	46	0	1	0	0	0	0	0	5	1	7	0	1	2
Normal 26	50	2.00	23.0	6000	7	15	2	0	0	0	0	0	0	0	2	0	1	0	0	0
Normal 27	63	2.00	692.7	6384	0	0	24	0	0	0	0	0	0	0	0	0	1	0	1	0
Normal 28	51	2.40	341.2	18085	1	0	59	0	0	0	0	0	0	0	0	1	0	1	0	0
Normal 29	51	2.20	15.9	6000	13	7	1	1	1	0	0	0	1	2	0	0	0	0	0	0
Normal 30	51	1.80	82.1	16000	0	9	29	1	0	0	0	0	0	1	3	0	1	0	0	0
Normal 31	61	2.10	26.0	11427	7	24	17	0	0	0	0	0	2	3	3	0	4	0	1	0
Normal 32	46	3.50	24.6	12568	2	32	7	2	0	1	1	1	1	0	0	0	0	0	0	1
Normal 33	46	3.50	5.9	2628	5	2	4	0	0	3	0	0	0	0	0	0	0	0	0	0
Normal 34	46	3.50	9.4	6502	4	21	6	1	2	1	0	0	1	1	0	0	2	1	0	1
Normal 35	46	3.50	7.8	5580	7	11	11	2	2	3	2	0	1	0	0	1	3	0	0	0
Normal 36	46	3.50	4.2	1776	1	4	0	0	1	0	2	0	1	0	0	0	0	0	0	0
Normal 37	46	3.50	2.5	1044	1	2	7	2	0	0	0	0	0	0	0	0	0	0	0	0
Normal 38	46	3.50	7.0	5606	4	17	11	0	0	3	1	1	2	1	2	2	2	0	0	0
Normal 39	47	3.50	12.2	7638	19	7	7	2	0	2	0	0	0	2	0	0	0	0	0	1
Normal 40	47	3.50	7.6	5138	3	7	1	1	1	0	0	0	2	0	0	0	0	0	0	0
Normal 41	47	3.50	3.7	1536	1	4	4	0	0	0	1	0	0	0	0	0	1	0	0	0
Normal 42	54	3.50	46.9	12448	0	19	10	4	0	3	0	0	0	0	1	0	2	0	0	0
Normal 43	53	3.50	2.3	936	6	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 44	48	3.50	6.9	5052	0	0	0	0	0	0	0	0	2	0	1	0	1	0	0	0
Normal 45	48	3.50	3.7	1560	3	3	1	1	0	2	0	0	0	0	0	0	0	0	0	0

Supplementary Table S1: First patient cohort of EOC (n=26) and normal control (n=45) plasma samples. Patient age, volume of plasma processed, cfDNA concentration (ng/mL), estimated genomic equivalents assessed in DREAMing, and number of observed melt peaks with a given methylation density for each sample are reported.

								DREAMing Profile (peak counts for each % methylation density)															
Sample	Patient blood draw timepoint	Age	Volume plasma (mLs)	ng/mL	CA125 U/mL	subtype	genomic equivalents loaded	0%	7%	14%	21%	28%	35%	42%	49%	56%	63%	70%	77%	84%	91%	98%	100%
Ovarian 1	Patient 1, P001	53	1.00	2.45	297	serous	288	15	2	1	1	0	2	4	0	0	0	0	2	5	1	1	0
Ovarian 2	Patient 2, P001	73	1.00	17.86	22	serous	2071	10	5	1	0	0	3	1	1	0	1	4	0	5	2	1	1
Ovarian 3	Patient 3, P001	44	1.00	3.13		serous	339	10	4	2	0	0	0	0	0	0	0	1	1	0	1	0	2
Ovarian 4	Patient 4, P001	70	1.00	12.17	326	serous	1476	2	11	1	1	1	2	0	0	2	1	0	0	1	1	1	0
Ovarian 5	Patient 5, P001	58	1.00	2.34	2.5	nonserous	254	6	5	7	0	0	1	1	1	3	1	2	1	3	0	0	0
Ovarian 6	Patient 6, P001	58	1.00	0.75	27	nonserous	84	2	2	1	0	1	0	1	0	1	0	0	0	3	0	0	0
Ovarian 7	Patient 7, P001	61	1.00	3.39	5.6	nonserous	372	2	14	1	0	1	0	0	1	4	1	2	1	2	1	0	0
Ovarian 8	Patient 8, P001	56	1.00	0.34	137	serous	48	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Ovarian 9	Patient 1, P002	53	1.00	4.52	358	serous	461	2	11	3	3	1	1	1	0	2	0	1	1	5	1	2	0
Ovarian 10	Patient 2, P002	73	1.00	23.21	15	serous	2710	15	4	1	1	0	1	1	0	0	1	1	0	6	1	1	0
Ovarian 11	Patient 3, P002	44	1.00	4.08	153	serous	485	2	15	0	2	2	2	1	0	1	1	1	1	5	1	0	0
Ovarian 12	Patient 4, P002	70	1.00	7.74	315	serous	919	3	9	2	1	2	2	0	0	2	1	0	0	1	0	1	0
Ovarian 13	Patient 5, P002	58	1.00	0.41	3.6	nonserous	48	2	3	3	0	0	0	1	0	0	0	1	1	2	0	0	0
Ovarian 14	Patient 6, P002	58	1.00	0.32	34	nonserous	48	0	1	0	2	0	0	0	0	1	0	1	0	2	0	1	0
Ovarian 15	Patient 7, P002	61	1.00	5.24	5.2	nonserous	584	2	12	2	1	0	1	1	1	3	0	2	2	2	0	0	0
Ovarian 16	Patient 8, P002	56	1.00	1.96	40	serous	225	5	9	0	1	0	0	0	0	1	0	0	1	1	0	0	0
Ovarian 17	Patient 1, P003	53	1.00	4.38	182	serous	530	14	2	1	1	0	1	1	0	0	1	0	0	0	1	0	2
Ovarian 18	Patient 2, P003	73	1.00	4.69	14	serous	521	11	2	1	5	0	4	1	0	1	0	0	1	2	0	0	0
Ovarian 19	Patient 3, P003	44	1.00	0.38	191	serous	24	0	0	1	1	0	1	0	0	0	0	0	0	1	1	0	0
Ovarian 20	Patient 4, P003	70	1.00	7.34	180	serous	882	2	9	3	1	1	3	3	1	2	0	2	2	2	1	0	0
Ovarian 21	Patient 5, P003	58	1.00	7.03	3.7	nonserous	777	12	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Ovarian 22	Patient 6, P003	58	1.00	40.80	26	nonserous	4720	6	3	5	2	1	3	1	0	1	0	1	1	0	0	0	0
Ovarian 23	Patient 7, P003	61	1.00	7.14	5.1	nonserous	823	1	12	2	0	0	1	0	1	2	1	0	0	2	0	0	0
Ovarian 24	Patient 8, P003	56	1.00	0.32	51	serous	24	3	3	1	0	0	0	1	0	1	0	1	3	0	0	0	0
Normal 1		56	1.6	12.50			2317	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 2		53	1.4	10.76			1309	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 3		51	1.6	7.54			1397	8	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 4		51	1.6	3.75			579	8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 5		52	1.6	2.17			301	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 6		45	1.8	2.64			458	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 7		48	1.5	3.27			474	3	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 8		49	1.4	2.64			392	10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Normal 9		65	1.6	9.98			1851	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Normal 10		65	1.6	3.43			635	0	10	0	0	0	1	0	0	0	1	0	0	0	0	0	0
Normal 11		48	1.8	3.59			747	0	11	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Normal 12		63	1.5	3.53			614	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Supplementary Table S2: Second patient cohort of EOC (n=24) and normal control (n=12) plasma samples. Blood draw timepoint, patient age, volume of plasma processed, cfDNA concentration (ng/mL), blood concentration of measured cancer antigen 125 (CA125, U/mL), EOC subtype, estimated genomic equivalents assessed in DREAMing, and number of observed melt peaks with a given methylation density for each sample are reported.