

DNA Methylation and Transcriptomic Features are Preserved Throughout Disease Recurrence and Chemoresistance in High Grade Serous Ovarian Cancers

Nicole Gull

Cedars-Sinai Medical Center <https://orcid.org/0000-0002-1956-5648>

Michell Jones

Cedars-Sinai Medical Center

Pei-Chen Peng

Cedars-Sinai Medical Center

Simon Coetzee

Cedars-Sinai Medical Center

Tiago Silva

University of Miami

Jasmine Plummer

Cedars-Sinai Medical Center

Alberto Reyes

Cedars-Sinai Medical Center

Brian Davis

Cedars-Sinai Medical Center

Stephanie Chen

Cedars-Sinai Medical Center

Kate Lawrenson

Cedars-Sinai Medical Center

Jenny Lester

UCLA: University of California Los Angeles

Christine Walsh

Cedars-Sinai Medical Center

Bobbie Rimel

Cedars-Sinai Medical Center

Andrew Li

Cedars-Sinai Medical Center

Ilana Cass

Dartmouth-Hitchcock Medical Center

Yonatan Berg

Hebrew University Hadassah Medical School: The Hebrew University of Jerusalem School of Medicine

John Paul Govindavari

Cedars-Sinai Medical Center

Joanna Rutgers

Cedars-Sinai Medical Center

Benjamin Berman

Hebrew University Hadassah Medical School: The Hebrew University of Jerusalem School of Medicine

Beth Karlan

UCLA: University of California Los Angeles

Simon Gayther (✉ simon.gayther@cshs.org)

Cedars-Sinai Medical Center

Research Article

Keywords: High grade serous ovarian cancer, methylation, chemoresistance, epigenetics, whole genome bisulfite sequencing, computational methods, translational research

Posted Date: January 6th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1214611/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Little is known about the role of global DNA methylation in recurrence and chemoresistance of high grade serous ovarian cancer (HGSOC). We performed whole genome bisulfite sequencing and transcriptome sequencing in 62 primary and recurrent tumors from 28 patients with stage III/IV HGSOC, of which 11 patients carried germline, pathogenic *BRCA1* and/or *BRCA2* mutations.

Results

Landscapes of genome-wide methylation (on average 24.2 million CpGs per tumor) and transcriptomes in primary and recurrent tumors showed extensive heterogeneity between patients but were highly preserved in tumors from the same patient. We identified significant differences in the burden of differentially methylated regions (DMRs) in tumors from *BRCA1/2* compared to non-*BRCA1/2* carriers (mean 659 DMRs and 388 DMRs in paired comparisons respectively). We identified overexpression of immune pathways in *BRCA1/2* carriers compared to non-carriers, implicating an increased immune response in improved survival ($P=0.006$) in these *BRCA1/2* carriers.

Conclusions

These findings indicate methylome and gene expression programs established in the primary tumor are conserved throughout disease progression, even extensive chemotherapy treatment, and that changes in methylation and gene expression are unlikely to serve as drivers for chemoresistance in HGSOC.

Background

Studies of the methylome and transcriptome in cancer have shown that conserved alterations can disrupt the molecular pathways that precede the onset of neoplasia and contribute to the development and progression of cancer and chemoresistance. The development of large scale genomic analyses of cancers have established that DNA methylation and its role in regulating gene expression is a key component in the development of most solid tumors. Germline or somatic hypermethylation is often an alternative mechanism to pathogenic loss of function mutations caused by coding and splice site mutations, deletions and rearrangements that lead to allele specific gene deregulation. Hypermethylation silences genes that are critical components of genome integrity (e.g., DNA repair genes, cell cycle regulation genes and tumor suppressor genes) and may be an early event in the progression to cancer. Indeed, loss of expression of genes involved in cancer development occurs about 10 times more frequently by DNA hypermethylation of promoter CpG islands than through mutation of DNA (1–3).

Initial studies focused on the role of methylation within CpG islands of gene promoters as a mechanism to silence gene expression (4–6), identifying key driver events in known cancer associated pathways. For example, somatic changes in methylation at the *BRCA1* promoter, result in reduced expression of *BRCA1*, homologous repair deficiency (7, 8) and shorter survival (9) in primary cancers; and alterations in methylation in the promoters of oncogenes (*PIK3CA*), transcription factors (*WT1*) and driver genes (*TP53*) have also been implicated in ovarian cancer development (10). Together, these data support a role of methylation in the regulation of genes that are critical in tumor initiation and development. Methylation is a genome-wide phenomenon that also targets the promoters of non-coding RNAs (11, 12) and more distal regulatory elements such as enhancers (13, 14). Thus, DNA methylation can also contribute to aberrant gene expression by altering the activity of transcription factor binding sites within enhancers and critical networks of gene expression variation involved in disease pathogenesis. More recent studies have shown that a global loss of methylation occurs in cancers, likely as part of the mitotic clock, across broad regions of the genome, known as partially methylated domains (PMDs) (13, 15–18). These regions generally harbor genes expressed at low levels and account for the bulk of methylation changes that occur in cancer (13, 15, 16, 18, 19).

Array based methods have become commonplace tools to evaluate the contribution of methylation to tumor pathogenesis. However, even the latest arrays measure the methylation status of only a small fraction of the nearly 30 million known CpGs throughout the genome (e.g., the latest iteration, the Illumina MethylationEPIC BeadChip array covers 850,000 CpG sites). Recent advances in whole genome bisulfite sequencing (WGBS) for methylation profiling provide single-base resolution, expanding our ability to identify functionally relevant DNA methylation regions on the basis of transcriptional regulation. WGBS analyses have not yet been performed in substantially large numbers of tumors, but the method is already providing novel insights into the role of methylation in cancer.

The contribution of DNA methylation to tumor recurrence and chemoresistance in ovarian cancer is poorly understood. Despite early indications that demethylating agents may be effective treatments for high grade serous ovarian cancer (HGSOC) (20) recent clinical trials of hypomethylating agents, including guadecitabine, have yielded little improvement in survival or resensitization to platinum-based therapies (21, 22). HGSOC is the most common and lethal histotype of ovarian cancer. About 70% of affected women are diagnosed with advanced stage disease (stages III/IV) and of these women <30% will survive more than five years. Patients are treated with maximal debulking surgery followed by combination chemotherapy with platinum. Typically, patients initially respond well to treatment, but usually relapse with recurrent and eventually chemoresistant disease (23). Between 15% and 25% of tumors are classified with primary resistance (24). This tends to occur in patients with homologous recombination proficient tumors and/or amplification of the *CCNE1* locus at 19q12 (7, 8). Nearly a third of all HGSOC cases have germline or somatic alterations in the *BRCA1* or *BRCA2* genes (7, 8), which result in DNA double strand break repair deficiency and an accumulation of DNA damage as tumors develop. The goals of this study were: (1) to establish the underlying role of DNA methylation in the recurrence and chemoresistance of HGSOC; and (2) to identify the role of DNA methylation in the development of HGSOC in women with and without germline defects in *BRCA1* and *BRCA2*.

Results

Whole genome bisulfite sequencing (WGBS) and transcriptomic profiling in matched primary and recurrent high grade serous ovarian cancer

We used WGBS to perform whole genome methylation profiling and RNA sequencing (RNA-seq) for whole transcriptome profiling of 62 fresh frozen matched primary and recurrent tumor tissues from 28 women diagnosed with stage III/IV HGSOC. Clinical features of the patients and their tumors are given in **Additional File 1, Supplementary Table 1** and illustrated in Figure 1A and B. Eleven of these patients carried a pathogenic germline mutation in the *BRCA1* and/or *BRCA2* genes. The remaining 17 patients were confirmed non-*BRCA1/2* germline mutation carriers (**Additional File 1, Supplementary Table 1**). All patients received similar first line treatments comprising optimal debulking surgery followed by combination chemotherapy with a platinum agent. Time to first recurrence and median survival times were significantly greater in *BRCA1/2* carriers compared to *BRCA1/2* non-carriers (2768 days vs 1678 days respectively, P-value=0.0056) (Figure 1B, 1C).

For each primary and recurrent tumor specimen, we generated a bisulfite converted library and sequenced to a depth of $\geq 30\times$, generating approximately 400 million read pairs per library, with a bisulfite conversion rate greater than 99%. After removing CpGs covered by fewer than 5 reads, we obtained on average 24.2 million CpGs per tissue specimen (range 13.1 - 26.5 million), with an average of 24.6 million CpGs in primary tumors (15.9 - 26.5 million) and 23.8 million CpGs in recurrent tumors (13.1 - 26.4 million). RNA was extracted from the same tissue samples and sequenced. We generated a mean of 335 million reads from each library. Samples with less than 90 percent of reads mapped to the correct strand of the reference genome were excluded. After removing very low expressed transcripts (<1 transcript per million (TPM)) and transcripts in 'blacklist' regions (25), we obtained 91,411 transcripts from 33,969 genes, coding and non-coding (Figure 1D).

Hypervariability in the landscape of partially methylated domains (PMDs) drives tumor heterogeneity between patients

Genome-wide maps of all CpG sites using supervised (Figure 2A) and unsupervised analysis (**Additional File 2, Supplementary Figure 1**) indicated widespread heterogeneity between tumors from different patients. Unsupervised clustering in each of the four sample groups (primary, recurrent, *BRCA1/2* carrier and *BRCA1/2* non-carrier) showed that methylation profiles in primary and recurrent tumors consistently clustered according to patient status rather than their sample group classifier, except that methylation profiles also clustered by *BRCA1/2* mutation status (**Additional File 2, Supplementary Figure 2A-D**).

We attributed this heterogeneity to a global loss of methylation within large genomic blocks (partially methylated domains or PMDs) (Figure 3A)(18, 26, 27). The fraction of the genome covered by PMDs varied from 1-58% in this set of tumors, with an average of 29% of the genome covered by PMDs per tumor (**Additional File 1, Supplementary Table 2**). There was no consistent pattern in the PMD architecture across this series of tumors. The most frequently observed PMDs were detected in 56/62 tumors, representing less than 0.03% of PMDs characterized in this set of tumors (Figure 3B). We identified a set of 'common' PMDs that were shared in more than 9 tumor specimens (ovcaPMDs), determined by the first inflection point of the bimodal distribution seen by plotting PMD frequency (Figure 3B; **Additional File 1, Supplementary Table 3**). Principal Component (PC) analysis using the top 10,000 most variable CpGs genome-wide (Figure 3C, **left**) showed that PC1 accounted for 48% of the total variance.

Hypervariable methylation patterns due to PMDs can be reproduced using only methylation values from soloWCGW CpGs, which are flanked by an A or T on both sides (palindromic) and reside alone within a window of +/- 35bp (18, 26). Thus, the average soloWCGW CpG

methylation was calculated to represent the PMD signature within a maximal set of PMDs that included both cell-type 'invariant' common PMDs (26) and ovcaPMDs. PC1 was highly correlated with soloWCGW PMD methylation (Pearson correlation = 0.74, P-value = 4.5×10^{-12} , Figure 3C, **left**). When we removed CpGs located within common+ovcaPMDs and re-performed PC analysis the variance accounted for by PC1 was reduced from 48–10%, and the overall association with soloWCGW PMD methylation was greatly reduced (Pearson correlation = 0.31, P-value = 0.01, Figure 3C, **right**), demonstrating that methylation in PMD regions contributes significantly to the heterogeneity across these tumors.

We also examined the effects of PMD masking by performing pairwise comparisons of soloWCGW methylation difference and Euclidean distance (calculated based on the 10,000 most variable CpGs) between each tumor (Figure 3D). The correlation between delta soloWCGW and pairwise Euclidean distance was significantly attenuated after PMD masking (Pearson correlation=0.65 and Pearson correlation = 0.17, respectively), reinforcing the observation that hypervariability at PMDs across the cohort contributes significantly to the observed heterogeneity.

CpG islands (CGIs) located within PMDs were hypermethylated compared to CGIs outside PMDs (**Additional File 2, Supplementary Figure 3A**). We extended this analysis to include other functional genomic elements. We observed increased levels of methylation at promoters but not CGI shores, gene bodies or intergenic regions within PMDs (**Additional File 2, Supplementary Figure 3B**). Whole transcriptomic profiling in the same tumor specimens was used to evaluate correlations for genes within common+ovcaPMDs. Genes within these PMDs were expressed at a significantly lower level than genes outside PMDs, with the lowest expression for genes inside the most common PMDs (P-value $< 2 \times 10^{-16}$, linear regression; Figure 4A, left panel). Also, there was more variability in the expression of genes within PMDs than for genes outside PMDs (P-value $< 2 \times 10^{-16}$, linear regression; Figure 4A, right panel). These findings were similar in tumors from different sample groups (*BRCA1/2* carrier and non-carriers, and from primary and recurrent tumors) (**Additional File 2, Supplementary Figure 4A, B**) and consistent with previously reported relationships between methylation and gene expression in PMDs (18).

We next examined the frequency of known or predicted tumor suppressor genes (TSGs) (28) located in common+ovcaPMDs. Seventeen percent of known TSGs (41/248) were located in these PMDs, of which only 17 TSGs (7%) are located in low-frequency PMDs (hypergeometric test P-value = 4.17×10^{-17} ; Figure 4B). TSGs known to be associated with ovarian cancer development were excluded from PMDs (P-value = 0.002). Seventy-one percent of the cancer genome atlas project (TCGA) signature genes (P-value = 2.75×10^{-18}) that define the four molecular subtypes of HGSOC (8) and 73% of pan-cancer oncogenes (28) were located outside PMDs (P-value = 1.94×10^{-9} ; Figure 4B). Taken together, these findings suggest that PMDs are depleted at genes important in cell identity and function. Gene Set Enrichment Analysis (GSEA) for genes located within PMDs shows that these genes are enriched in the RIG-I-like receptor signaling and ERBB signaling pathways (**Additional File 1, Supplementary Table 4**), similar to previously reported findings in breast cancer (18) and human neuron cells (29).

We masked the genome for PMD methylation to remove the influence of highly variable PMDs and then reanalyzed the data. Hierarchical clustering and PC analysis for the 10,000 most variable CpGs showed that recurrent tumors continued to cluster most closely with the matched primary tumor from the same patient rather than by primary, recurrent or *BRCA1/2* mutation status (**Additional File 2, Supplementary Figure 5A, 5B**). None of the 10,000 most variable CpGs are known methylation QTLs in normal or cancer tissues, indicating this clustering was not driven by germline genetic variation (30, 31). To quantify these clusters we calculated the Euclidean distances from each primary and recurrent tumor pair from the same patient (intra-patient distances) and the distances of all pairs from different patients (inter-patient distances). Intra-patient distances were significantly smaller than inter-patient differences in both *BRCA1/2* carriers and non-carriers (P-values = 7.16×10^{-7} and 1.41×10^{-3} respectively; Figure 4C). For further validation of this finding, we evaluated heterogeneity associated with whole transcriptome data. Similarly, intra-patient Euclidean distances were significantly shorter than inter-patient distances in both *BRCA1/2* carriers and non-carriers (P-values = 9.67×10^{-5} and 6.70×10^{-4} respectively; Figure 4C). These observations are unlikely to be the result of residual PMD hypomethylation differences that remain after masking, but more the epigenetic landscape of recurrent tumors defined by inter-patient heterogeneity.

HGSOCs maintain methylation and gene expression programs in primary tumors progressing to chemoresistance

We used metilene to perform paired analysis of the primary and recurrent tumors from each patient (32) to identify changes in methylation associated with recurrence after a diagnosis of primary HGSOC. We identified 15,082 DMRs with >10% change in methylation and $Q < 0.1$ across all primary and recurrent tumor pairs, but found no DMRs meeting a genome-wide corrected significance threshold of $Q < 0.05$. We found significantly more DMRs (>10% change, $Q < 0.1$) in primary versus recurrent tumors from *BRCA1/2* non-carriers (11,205 DMRs in 16 tumor pairs, average of 659 DMRs per case comparison) compared to tumors from *BRCA1/2* carriers (3,877 significant DMRs in 11 tumor pairs, average of 388 DMRs per case comparison) (P-value = 0.004). We restricted these analyses to 1,785 DMRs that were

shared between primary and recurrent tumors from 2 or more patients (**Additional File 1, Supplementary Table 5**). We calculated the change in methylation at each CpG site within these DMRs between primary and recurrent tumors from each patient. Of these DMRs, 558/1785 (31%) were discordantly methylated (in other words, the same region is observed as hyper- and hypo-methylated in different patients) (**Additional File 2, Supplementary Figure 6**); but hierarchical clustering analysis failed to identify any clinical or molecular features that correlate with consistent methylation changes at these regions. Taken together, these data indicate that the methylation landscapes of recurrent HGSOCs are largely conserved after chemotherapy and disease recurrence, and do not acquire common somatic methylation changes that may be drivers of chemoresistance and tumor recurrence.

Whole transcriptome data identified 99 genes that were differentially expressed between all primary and recurrent tumors (adjusted P-value <0.05), of which 37 genes had lower expression and 62 genes higher expression in recurrent compared to primary tumors (**Additional File 1, Supplementary Table 6**). Differential expression for twenty of these genes was directionally consistent with differential changes in methylation (**Additional File 1, Supplementary Table 7**). We compared the DEGs we identified with 1,836 known/characterized tumor suppressor genes and oncogenes associated with pan-cancer development and molecular signature genes for different HGSOC subtypes described by TCGA (8). Two genes– the oncogenes *CCNE1* and *DDX5* – were differentially expressed between primary and recurrent tumors. *CCNE1*, which is significantly amplified and highly expressed in aggressive HGSOC cases with poor clinical outcome (33), was overexpressed in recurrent tumors ($P=3.65 \times 10^{-6}$). *DDX5* was overexpressed in recurrent tumors and located within 10 genes of a DMR that is hypomethylated, although at ~270kb from the gene (**Additional File 1, Supplementary Table 7**).

Hypomethylation increases expression of immune-related genes and identifies putative drivers of disease recurrence in HGSOCs from BRCA1/2 carriers

We identified 135 significant DMRs in primary-recurrent HGSOCs between *BRCA1/2* carriers and non-carriers (Q-value < 0.05)(Figure 5A; **Additional File 1, Supplementary Table 8**), with a trend for hypermethylation at these DMRs in tumors from *BRCA1/2* non-carriers and genome-wide hypermethylation in PMD regions in primary tumors from *BRCA1/2* non-carriers (P-value = 0.0011) (**Additional File 2, Supplementary Figure 7**). PCA analysis also suggested a trend for tumors to cluster together based on *BRCA1/2* carrier status (Figure 5B). Annotation of DMRs hypermethylated in *BRCA1/2* non-carrier tumors showed these regions were depleted in CpG islands near transcription start sites and active regulatory regions marked by H3K27ac in the secretory epithelial fallopian tube cell line FT246 (a model for the precursor cell type for HGSOC) indicated by at least a 15% reduction in methylation over background (**Additional File 2, Supplementary Figure 8**). This depletion indicates that while heterogenous, differential methylation between *BRCA1/2* carrier and non-carrier tumors is focused at active regulatory regions in relevant cell types and CpG islands near transcription start sites, which play an important role in the regulation of gene expression.

We performed a replication analysis of the 135 DMRs using previously published array-based differential probe methylation data generated using Illumina 450k array analysis of tumors from 20 *BRCA1/2* carriers and 60 non-carriers (7) ($q < 0.05$, **Additional File 1, Supplementary Table 9**). For this analysis, we masked probe locations that overlapped common+ovca PMDs, reducing total probe count on the array to 310,968 probes. These probes were intersected with our DMRs, and overlapping probes were used to identify DMRs between *BRCA1/2* carriers and non-carriers using the tool bumphunter (34), which requires two probes within 300bp to have a significant change in methylation in the same direction of one another to be considered a DMR candidate region. After removing probes based on these criteria that did not overlap DMRs, replication data were available for 31/135 DMRs. We did not identify significant differentially methylated regions using bumphunter at any of these 31 DMRs. A similar replication analysis was not possible using TCGA dataset (613 HGSOCs including 52 *BRCA1/2* and 561 *non-BRCA1/2* tumors) because these tumors were analyzed using the Illumina 27K methylation array for which there is very low concordance (one CpG probe, cg21557231) with our WGBS data.

PCA analysis based on differential gene expression (DEG) analysis also identified clusters associated with *BRCA1/2* mutation status (Figure 5C). We identified 3,341 DEGs in tumors from *BRCA1/2* carriers compared to non-carriers (adjusted P-value <0.05) (**Additional File 1, Supplementary Table 10**), of which 1,760 genes were up-regulated and 1,581 genes were down-regulated (Figure 5D). This is consistent with DMR analysis where we found a greater proportion of hypomethylated CpGs in tumors from *BRCA1/2* carriers compared to non-carriers. Up-regulated genes in *BRCA1/2* associated tumors were significantly enriched in immune related pathways (Q<0.05) including autoimmune diseases, infection response and antigen processing and presentation (Figure 5E), even though there were no clear differences in immune cell infiltration in *BRCA1/2* tumors compared to non-*BRCA1/2* tumors (consensusTME scores P-value = 0.97) (**Additional File 2, Supplementary Figure 9**). Down-regulated genes in tumors from *BRCA1/2* carriers were most significantly enriched in pathways that maintain stemness and cell differentiation, including the hippo signaling pathway (adjusted P-value <0.05) (Figure 5E).

We connected changes in the methylation of regulatory elements to changes in gene expression. One thousand seven hundred and eighty two genes were located within 10 genes up and down-stream of the 135 DMRs (**Additional File 1, Supplementary Table 11**), including 94 gene promoters that were within 2kb of a DMR. We compared this list with the 3,341 DEGs identified in tumors of *BRCA1/2* carriers and non-carriers and identified 68 instances (includes some duplicate genes) where DMRs showed either positive or inverse correlation between methylation and expression of the nearby gene (**Additional File 1, Supplementary Table 11**). This corresponded to 37 unique regions. These DMRs varied in length from 100bp to 8kb with the number of promoters within 2 kb of a single DMR ranging between one and six. Twenty-five unique genes were hypermethylated and down-regulated in *BRCA1/2* carriers while 41 unique genes hypomethylated and up-regulated in tumors from *BRCA1/2* carriers. (**Additional File 1, Supplementary Table 11**). We adapted the software tool ELMER (35, 36) to correlate methylation values in DMRs with the expression of nearby genes, identifying three genes - *FGF18*, *CDK2AP1*, and *NAGLU* - where methylation and expression were positively correlated ($q < 0.05$, $r = 0.46$, 0.44 , and 0.48 respectively) (Figure 5F). Positive correlation between methylation and expression can indicate that a gene's function is affected by gene body methylation or possibly methylation of a nearby insulator (37), although paradoxically increased expression may also arise from a fully methylated promoter (38).

Discussion

Whole genome bisulfite sequencing (WGBS) catalogs a comprehensive genome-wide epigenetic landscape, with coverage of CpGs at several orders of magnitude greater than the array-based approaches that have largely used up until now to characterize the genome-wide methylation status of tumors. In this study, WGBS analysis reported on the methylation status of, on average, 24.6 million CpG sites per tumor analyzed. This contrasts with methylation arrays that interrogate highly selected CpG sites, of which the most commonly used have been the Illumina 27K, 450K, and EPIC (850K) arrays that evaluate 0.1%, 1.5% and 3.0% of CpGs in the genome, respectively.

The current study is the first to use WGBS instead of an array to comprehensively map CpG methylation and the transcriptome in matched primary HGSOCs and tumor recurrences arising post chemotherapy in the same patient. It is also the first study to comprehensively map PMDs genome-wide in HGSOCs using WGBS. The data are consistent with studies of the PMD architecture of other cancers and tissues (13, 17, 18, 26, 27). We identified a common set of PMDs, encompassing 15% of the genome that appear to show specificity to HGSOC. Within these PMDs, CpG islands and other functional genomic elements were highly methylated and genes were expressed at lower levels compared to those located outside PMDs (13, 17, 18). While the distribution of PMDs in primary and recurrent tumors was heterogeneous, there was a strong and highly statistically significant enrichment for genes involved in cancer development, and more specifically genes that are differentially expressed and used to stratify HGSOCs into different molecular subtypes, located outside PMDs. We postulate that the critical nature of these genes in both normal cell function and tumor development requires these genes to be active, both spatially and temporally in their differentiation into specific tumor phenotypes. PMD hypomethylation was a central feature of the variation in methylation we observed across our samples and by masking these regions we were able to find differences in methylation between *BRCA1/2* carriers and non-carriers. Without PMD masking, these differences were not apparent.

Given that our study is the first WGBS analysis of primary HGSOCs and of their recurrences, replicating our results in an independent cohort presents significant challenges. The largest genomic studies of HGSOC has been performed by TCGA, which includes methylation analysis using the 27K CpG array of 489 primary HGSOCs compared to 8 full thickness fallopian tube tissues. This analysis identified 168 genes that were epigenetically silenced in HGSOC. Twenty nine of these genes showed significantly reduced RNA expression in non-*BRCA1/2* tumors analyzed in our study, including four of the fifteen top ranked genes (*AMT*, *LDHD*, *CFTR* and *BANK1*) that we identified in a comparison of gene expression HGSOCs from *BRCA1/2* carriers compared non-carriers. Our WGBS analysis also identified some differentially methylated genes that have been identified in HGSOCs in other studies, notably *HOXA9* which was found to be methylated in up to 95% of ovarian cancers in a study of 80 primary tumors from Montavon and colleagues (39). *MYO18B* inactivation has also been reported in chemoresistant HGSOCs; this gene was significantly downregulated in tumors from *BRCA1/2* carriers in our analysis (40).

Previous studies have compared the molecular features of primary ovarian cancers (including methylation) with metastatic tumors or cells from ascitic fluid to identify molecular biomarkers and novel therapeutic targets for chemoresistant disease (7, 8, 41–43). Generally, these studies suggest there is an accumulation of somatic mutations as tumors metastasize and recur after chemotherapy (7, 42), although no frequently somatically altered genes associated specifically with tumor metastasis/recurrence have yet been identified. While we did not perform whole genome sequencing of the same tumor specimen in the current study, our data from whole genome methylation and transcriptomic sequencing are largely consistent with previous findings. The most prevalent genes that were differentially altered in recurrent tumors were the known oncogene *DDX5*, which has been shown to be frequently amplified in breast cancers (44), and *CCNE1*, which is frequently amplified in HGSOC (33). Otherwise, we found few molecular changes that were specific to tumor recurrences. Methylation and gene expression signatures were highly preserved in recurrences relative to the primary tumor from the same individual, but were significantly different between tumors from different individuals. There was little evidence of an accumulation of additional and

novel methylation and transcriptomic changes in the recurrent tumors that could be attributed to chemoresistance. This result is similar to a previous study of HGSOC using a methylation array that interrogated 91 patients with primary HGSOC and 6 non-matched patients with recurrent HGSOC (45). This is notable as it indicates that the genomic changes required to promote chemo-resistance may be established early in the evolution of the primary tumor and persist to dominate the clonal populations of both the primary and subsequent recurrent tumors. This is consistent with a previous study of somatic mutations identified in multiple primary and metastatic samples from seven ovarian cancers, which found complex patterns of both monoclonal and polyclonal seeding of metastatic sites and predicted a lack of selective pressures after treatment with combination chemotherapy (46). This may indicate the presence of clones of chemoresistant disease at the time of primary tumor diagnosis. Thus, while the bulk of disease would respond to platinum-based therapies, a proportion of tumor cells may persist through chemotherapy to seed recurrent tumor growth within the peritoneal cavity.

Our WGBS analysis indicates that there are striking differences in the methylation profiles between patients with and without germline *BRCA1/2* mutations. The data suggest that hypermethylation is a feature of non-*BRCA1/2* associated tumors. This result was consistent within PMDs, specifically at soloWCGW CpGs and adds to the growing body of evidence that indicates non-*BRCA1/2* associated ovarian cancers develop along different molecular pathways to ovarian cancers from *BRCA1/2* carriers. Recent studies have suggested that foldback inversions may be drivers of HGSOC development in *BRCA1/2* non-carriers resulting in unique mutational processes that do not correlate with any of the different molecular subtypes described for HGSOC by TCGA (8, 23). Our data support the findings of studies that have identified several notable genes that are significantly differentially expressed in non-*BRCA1/2* compared to *BRCA1/2* ovarian cancers including: *EIF3CL* which regulates a cluster of metastasis-promoting genes via *STAT3* and acts as a mediator of immune cell evasion (47) and *CFTR* overexpression (also reported by TCGA) which is known to increase cell invasion, proliferation and adhesion in ovarian cancers (48), and is highly expressed in the fallopian tube secretory epithelial cells from which many HGSOCs arise (49).

Conclusions

Our study describes the first comprehensive analysis of paired methylation landscapes generated by WGBS in HGSOCs and their recurrences after chemotherapy, and the first comparison in patients with and without germline *BRCA1* and *BRCA2* mutations. This study highlights the molecular heterogeneity that exists, and has been recently demonstrated using single-cell data (50, 51) amongst HGSOCs and provides evidence that this heterogeneity extends to chemoresistant, recurrent disease. We have demonstrated that there is an absence of common methylation signatures or specific methylation biomarkers that would indicate common mechanisms and underlying biology associated with disease recurrence or chemoresistance. This observation was replicated using whole transcriptome profiling in the same primary-recurrent tissue specimens. Taken together, these data suggest that the methylation and transcriptomic changes required to survive first line chemotherapy and seed recurrent tumors may already be present in the primary tumor, rather than induced as a result of exposure to chemotherapy. The most significant methylation and/or transcriptomic variations were observed when we compared primary tumors with and without *BRCA1* or *BRCA2* mutations. The improved survival and disease free interval in patients with *BRCA1* or *BRCA2* mutations has been attributed to their improved response to platinum based chemotherapy, and we have identified extensive differences in the methylome and transcriptome between these groups that likely contribute to these differences.

Methods

Experimental Design

Cohort description

Fresh-frozen matched primary and recurrent high-grade serous ovarian cancer specimens and DNA from twenty-seven consented women diagnosed with high grade serous adenocarcinoma were included, as well as one additional primary tumor specimen with no matched recurrent tumor (28 patients total). Specimens were all identified in the Cedars-Sinai Medical Center Women's Cancer Program Biorepository (IRB #0901). 11 of these women had deleterious *BRCA1* and/or *BRCA2* germline mutations and 17 did not harbor any known high or moderate risk mutations for HGSOC (*BRCA1*, *BRCA2*, *RAD51C*, *RAD51D*, *BRIP1* and *FANCM*) identified in clinical genetic testing. All patients were diagnosed with stage III or stage IV disease and underwent primary optimal surgical cytoreduction (to less than 1cm residual disease) prior to administration of combination chemotherapy with platinum and taxane between the years of 1990 and 2014. For each patient, detailed clinical data were available including clinical genetic testing results, dates of original diagnosis and each subsequent recurrence, treatments administered throughout their disease course, operative and pathology reports, and other clinicopathologic variables including other cancer diagnoses and comorbidities. These details can be found in **Additional File 1, Supplementary Table 1**.

Specimen acquisition and preparation

Fresh frozen tumors were embedded in optimal cutting temperature (OCT) compound, bisected and mounted and two slides were made for hematoxylin and eosin (H&E) staining. All slides were reviewed by a single pathologist to identify regions enriched for epithelial carcinoma (avoiding tumor stroma), which are then collected in a single punch of approximately 50mg collected on dry ice. Each punch was divided into three pieces, two of which were used for genomic DNA (gDNA) extraction using the Machery-Nagel Nucleospin DNA Kit, and the third for RNA extraction using the Machery-Nagel Nucleospin RNA Kit. DNA samples were assayed for quality using the Qubit (Thermo Fisher Sci, CA) to measure the content of double stranded DNA and by running 1ul gDNA on a 1.5% agarose gel at 100V for 1 hour to confirm no fragmentation of material has occurred during the extraction. RNA was extracted using an isopropyl-alcohol:chloroform approach following the standard operating protocol published by the Prostate Cancer Biorepository Network(52). RNA was quantified on the Qubit in RNA mode to measure the amount of high quality dsRNA within the sample, and then on the Agilent Bioanalyzer, where an RNA Integrity Number (RIN) score is generated, reflecting the quality (by concentration and fragment size) of the samples. Germline DNA was extracted from whole blood drawn at the time of debulking surgery after diagnosis with HGSOc. DNA was extracted with the Qiagen DNEasy Blood & Tissue Kit (Qiagen, Germantown, MD, USA) and quantitated with the Quant-IT dsDNA Broad Range kit on a Qubit (Thermo Fisher, Waltham, MA, USA).

Method Details

Whole genome bisulfite sequencing (WGBS)

Our workflow for WGBS required a minimum 300ng of high quality gDNA, which was sheared to approximately 175-200 bp using a Covaris sonicator, and bisulfite converted using the EZ DNA Methylation-Lightning Kit (Zymo). Libraries were constructed using the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences, MI), and amplified using no more than 6 cycles of PCR. Libraries were sequenced to at least 30x coverage (on average each base is sequenced thirty times) on the Illumina HiSeq4000 in 150bp paired end mode. This approach generated approximately 400 million read pairs per library, with a bisulfite conversion rate greater than 99%.

RNA-Seq library preparation and sequencing

RNA was extracted using the protocol published online by the Prostate Cancer Biorepository network (SOP#006), where frozen tissue was stored at -80C until extraction. 1mL of trizol was added to the tissue in a 1.5mL eppendorf tube and incubated for 5 mins at 15-30C to dissociate nucleoprotein complexes. Next 200ul of chloroform was added and tubes capped and mixed vigorously for 15 secs then incubated at room temperature for 3 mins. Samples were then centrifuged for 15 mins at 4C at 12,000g. The top, aqueous phase was removed to a fresh, sterile 1.5ml tube and mixed with 500ul of isopropyl alcohol to precipitate the RNA. Samples were then incubated at room temperature for 10 mins and then centrifuged for 10min at 4C at 12,000g. A pellet was visible in high yield samples, and supernatant was removed (and discarded), leaving the pellet untouched. The pellet was washed with 1mL 75% ethanol and mixed by vortexing and then centrifuging at 7,500g for 5min at 4C. The supernatant was removed (leaving the pellet untouched) and the pellet allowed to air dry before being dissolved in 200ul RNase-free water and incubated at 55C for 10 mins. Sample concentration was measured using the Qubit RNA Broad Range kit and sample quality was measured using the Agilent BioAnalyzer 2100. Sequencing libraries were prepared by adding 1ug of RNA to the TruSeq Stranded Total RNA Kit with Ribo- and Mito- depletion following the TruSeq standard protocol with 15 cycles of PCR. Libraries were quantified using Qubit RNA Broad Range kit and pooled before being run on one lane of a HiSeq2000 to collect ~1M reads per library for quality control. PCR duplication rate was estimated in this low coverage sequencing run in 150bp paired end mode, and library complexity was estimated using PreSeq. Based on the complexity measured in this low coverage sequencing experiment we estimated the maximal coverage that would continue to provide informative measurement of transcripts in the library was ~350M reads. Each library was then pooled and this pool was sequenced in 2x150bp mode on an Illumina Novaseq 6000, and we generated ~335 million reads from each library. Our data analysis workflow for RNA-Seq has been developed specifically to improve gene feature identification and measurement in archived frozen tissue samples, which can perform poorly using standardized workflows.

Statistical Analysis

WGBS data processing

WGBS reads were aligned to the human reference genome (build GRCh38) using BISCUIT (53). Duplicate reads were marked using Picard Tools (54). Methylation rates were called using BISCUIT. CpGs with fewer than 5 reads of coverage were excluded from further analysis. Adapter sequences were trimmed using TrimGalore (55), using default parameters for Illumina sequencing platforms. Quality control was performed using PicardTools as well as MultiQC (56). Bisulfite non-conversion was checked using the Biscuit QC module in MultiQC.

Principal Component (PC) analysis was performed on CpGs with coverage ≥ 10 and the top 10,000 most variable CpGs were included in the identification of the top 10 PCs using the `prcomp` function from the `stats` package in R (57).

Calling partially methylated domains (PMDs)

To call Partially Methylated Domains (PMDs), we first divided the genome of each sample into non-overlapping 100kb bins, and took the average of all solo-WCGW CpGs within each bin, using the solo-WCGW definition from Zhou et al (26). We then converted the methylation averages to M-values ($M_i = \log_2(\text{Beta}_i / (1 - \text{Beta}_i))$) (58), and fit M-values to a 3-component Gaussian Mixture Model (GMM) using the `mclust` R package (59). Based on its mean, we assigned the three components labels of low, intermediate, and high. Each bin was labeled as a PMD if the probability of being in the high bin was less than 0.01, and multiple consecutive PMD bins were merged into a single PMD call. Common ovarian cancer PMDs (ovcaPMDs) were defined as regions identified from `PMDetect` in more than 9 of our samples. Common ovca PMDs were combined with PMDs from other cancer cell types (26). Solo-WCGW scores were calculated by averaging the methylation of solo-WCGW CpGs (using the definition from Zhou et al (26)) within the combined ovca+commonPMD set. Taken together these PMDs spanned 69.57% of the genome, comprising 14.96% of the genome spanned by ovcaPMDs and 54.6% of the genome spanned by the common PMD set.

Identifying variable CpGs that overlap known mQTLs

Methylation quantitative trait loci (mQTL) from two publicly available datasets that represent known mQTLs in normal tissues (30) and cancer (31). mQTLs with P-value $< 5.0 \times 10^{-8}$ were included for intersection with the 10,000 most variable CpGs across the cohort after PMD masking. This included 4,045,382 unique CpGs that are mQTLs across normal tissues and 41,057 unique CpGs that are mQTLs across 23 cancer types.

Calling differentially methylated regions (DMRs)

PMDs were masked from each sample's BED file before conducting differentially methylated region (DMR) analysis. The Bioconductor package `dmrseq` (60) was utilized to identify DMRs between *BRCA1/2* carriers and *BRCA1/2* non-carriers using default settings. `Metilene` (32) was used to identify DMRs between matched primary and recurrent tumors from each patient using the following parameters: `-M 500, -m 5, p 0.1, -c 5`. Only DMRs that overlapped between two or more patients were retained. `Bedtools` (61) `merge` function was performed on all overlapping DMRs to merge regions within 250bp using parameter `-d 250`. Heatmaps were plotted using mean methylation across each identified DMR. Enrichment analysis of DMRs was conducted using `annotatr` (62) and `ChIPseeker` (63). We also included an enrichment analysis of several other genomic classifications, such as DNA methylation valleys, and PRC2 binding regions. In the package `annotatr`, we searched for enrichment in coding regions, intergenic regions, exons, introns, 3'UTR, CpG shores, shelves and islands. Using `ChIPseeker`, we searched for enrichment in the default regions: promoter, 5'UTR, 3'UTR, exons, introns, and intergenic regions. We refined promoter regions based on distance from transcription start site (TSS) and redefined CpG islands and CpG shores as non-TSS or TSS if the region was within 2kb of a TSS. Backgrounds for enrichment controlled for either DMR size only or DMR size plus CpG count. A DMR size only background was generated using `bedtools shuffle` (61,64). DMR size plus CpG count background was generated using an in-house developed script.

Replication of DMRs with publicly available datasets

Genome-wide methylation data for primary HGSOC tumors from 20 *BRCA1/2* carriers and 60 *BRCA1/2* non-carriers from the Illumina 450k array was downloaded from the Gene Expression Omnibus (GEO) at Accession GSE65821 (7). DMR calling was performed using `minfi` (34), which requires two probes within 300bp to have $> 20\%$ change in beta value in the same direction. Due to the promoter-focused design of the array only 31 of our 135 DMRs identified in the comparison of *BRCA1/2* carriers and non-carriers intersected two or more 450k probes within 300bp. Many of our DMRs were within enhancers that were intergenic or in intron 1. To include local regions (primarily promoters) with sufficient probe coverage such that `minfi` could identify DMRs should they be present, we extended the boundary of each of the 135 DMRs by 1kb. 70 of our 135 DMRs now met the basic criteria for `minfi` to identify a DMR at that location, should one exist. DMR analysis of all probes within this set of coordinates did not produce any replication.

RNA-Seq quantification and statistical analysis

Reads within each fastq file were first trimmed using `TrimGalore` to remove low quality bases and sample barcodes, retaining reads 75bp or longer. Each transcriptome is then aligned to hg38 and the Gencodev29 primary assembly (64). Genes are quantified with `RSEM` (65) and `Kallisto` (66). Sample-specific gene models were generated using alignments produced with `STAR` two pass mapping and `Stringtie` (67). Gene expression values were shown as normalized variance stabilizing transformation (vst) counts. To measure RNA abundance, we

first obtained BAM alignment quality metrics using Picard (<http://broadinstitute.github.io/picard>). Samples with less than 90 percent of reads mapped to the correct strand of the reference genome (PCT_CORRECT_STRAND_READS) were omitted. Patients whose primary and recurrent tumors both passed this quality control were retained (n=50) (**Additional File 1, Supplementary Table 12**). Read counts were quantified using the R package Salmon (68) at transcript level and reads were mapped to Genecode Release 29 (GRCh38) comprehensive gene annotations by R package 'tximport' (69). To filter out potential artifacts and very low expressed transcripts we retained transcripts with length greater than 300 bp, TPM (Transcripts Per Million) value greater than 0.05 and isoform percentage greater than 1%. Transcripts in blacklist regions (25) were also filtered out. We retained transcripts expressed in more than 5 samples, which resulted in 91,411 transcripts from 33,969 genes. Tumor purity was estimated by the degree of heterogeneity of the tumor microenvironment. We applied the R package 'consensusTME' (70) to estimate cell type specific enrichment scores based on TCGA ovarian cancer data, and confirmed there was low stromal content and generally low infiltration by immune cells across the samples (**Additional File 2, Supplementary Figure 9**).

Differentially expressed genes between the *BRCA1/2* carrier versus *BRCA1/2* non-carrier and primary versus recurrent tumors were detected by R package 'DESeq2' (71). The P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure. Since our experiment design has group-specific effects, comparisons between *BRCA* carrier status are made between patients, while comparisons between primary versus recurrent tumors are made within the patient. To control for confounding differences between the primary and recurrent tumors from patients we constructed a nested DESeq2 model with formula; ~purity + BRCAStatus + BRCAStatus:PatientID + BRCAStatus:isRecur, which has the main effect for BRCA status plus nested interactions with primary and recurrent status. To see whether the identified gene sets (e.g. genes inside PMDs or differentially expressed genes) show significant functional concordance, we performed Gene Set Enrichment Analysis (72) for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (73). We implemented enrichment analysis with R package 'clusterProfiler' (74). For each enrichment analysis, we set the number of permutations to 10,000 and reported enriched pathways with a Benjamini-Hochberg adjusted P-value less than 0.05.

Linking enhancers to target genes

Correlation between DMRs and gene expression was performed by comparing primary and recurrent tumors from each individual patient, for samples with matched WGBS and RNA-seq available for both primary and recurrent tumors (n = 27). Only DMRs that overlapped between two or more patients were included in this analysis. Using GENCODE28, DMRs regions >2kb from any TSS were annotated as "distal" and regions <2kb from TSS were annotated as "promoter". Distal regions were mapped to the closest genes (10 upstream and 10 downstream) and the promoter to the closest gene and the correlation between their expression and methylation measured, where average beta value of the DMR correlates (using Spearman test) to a change (positive or negative) in expression of the nearby genes. ELMER version 2.8.3 (36) was used to map the genes, and the correlation was performed to each link (DMR - gene) only using the samples in which the DMR was identified using the function cor.test in R. Links with a minimum P-value of 0.05 were retained.

Declarations

Ethics approval and consent to participate

Informed consent was obtained from all twenty-eight women included in this study, agreeing to provide fresh-frozen high-grade serous ovarian cancer specimens and DNA, as well as detailed clinical data including clinical genetic testing results, dates of original diagnosis and each subsequent recurrence, treatments administered throughout their disease course, operative and pathology reports, and other clinicopathologic variables including other cancer diagnoses and comorbidities. All women were diagnosed with high grade serous adenocarcinoma. Specimens were all identified in the Cedars-Sinai Medical Center Women's Cancer Program Biorepository under Institutional Review Board Protocol #0901.

Consent for publication

Not Applicable

Availability of data and materials

This study did not generate any unique reagents or materials. Project data will be made available prior to acceptance of the article. Code to generate most figures as well as unique code required for this study will be uploaded and maintained on Github by the first author, Nicole Gull: <https://github.com/nicoleversetwo/methylation-foo>

Competing Interests

The authors declare that they have no competing interests

Funding

Cedars-Sinai Medical Center Precision Healthy Initiative

NIH/NCI R01CA211575 (SAG)

U01CA184826 (BPB, SGC)

Samuel Oschin Comprehensive Cancer Institute Cedars-Sinai Medical Center Developmental Funds (BPB)

Ovarian Cancer Research Alliance Liz Tilberis Early Career Award, grant 599175 (KL)

Ovarian Cancer Research Alliance Program Project Development Award, grant 373356 (KL)

Research Scholar's Grant from the American Cancer Society, grant 134005 (KL)

National Cancer Institute, grant K99CA256519 (PCP)

Generous donation from the Barth Family (SAG)

Author Contributions

Study design: MRJ, BYK, SAG, KL, BPB

Sample collection: JL, CW, BJR, AJL, IC, BYK

Sample processing, data acquisition and management: MRJ, JTP, BDD, SC, KL, AR, JL, JPBG, JKLR

Data analysis and interpretation: NG, MRJ, PCP, SGC, TCS, YB, BPB, BYK, KL, SAG

Manuscript writing: NG, MRJ, PCP, KL, BYK, BPB, SAG

Decision to submit for publication: NG, MRJ, BYK, SAG

Acknowledgements

We would like to acknowledge the generous contributions made by the patients who participated in this study.

References

1. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet*. 2010 Sep 23;6(9):e1001134.
2. Wei J, Li G, Zhang J, Zhou Y, Dang S, Chen H, et al. Integrated analysis of genome-wide DNA methylation and gene expression profiles identifies potential novel biomarkers of rectal cancer. *Oncotarget*. 2016 Sep 20;7(38):62547–58.
3. Dilworth MP, Nieto T, Stockton JD, Whalley CM, Tee L, James JD, et al. Whole Genome Methylation Analysis of Nondysplastic Barrett Esophagus that Progresses to Invasive Cancer. *Ann Surg*. 2019;269(3):479–85.
4. Millar DS, Ow KK, Paul CL, Russell PJ, Molloy PL, Clark SJ. Detailed methylation analysis of the glutathione S-transferase pi (GSTP1) gene in prostate cancer. *Oncogene*. 1999 Feb 11;18(6):1313–24.
5. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst*. 2000 Apr 5;92(7):564–9.
6. Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, et al. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res*. 1995 Oct 15;55(20):4525–30.
7. Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature*. 2015 May 28;521(7553):489–94.
8. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011 Jun 29;474(7353):609–15.

9. Bolton KL, Chenevix-Trench G, Goh C, Sadetzki S, Ramus SJ, Karlan BY, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA*. 2012 Jan 25;307(4):382–90.
10. Feinberg AP, Koldobskiy MA, Göndör A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet*. 2016 May;17(5):284–99.
11. Klungland A, Dahl JA, Greggains G, Fedorcsak P, Filipczyk A. Reversible RNA modifications in meiosis and pluripotency. *Nat Methods*. 2016 Dec 29;14(1):18–22.
12. Patil DP, Chen C-K, Pickering BF, Chow A, Jackson C, Guttman M, et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*. 2016 Sep 15;537(7620):369–73.
13. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet*. 2012 Jan;44(1):40–6.
14. Aran D, Toperoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet*. 2011 Feb 15;20(4):670–80.
15. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011 Jun 26;43(8):768–75.
16. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res*. 2012 Feb;22(2):246–58.
17. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med*. 2014 Aug 26;6(8):61.
18. Brinkman AB, Nik-Zainal S, Simmer F, Rodríguez-González FG, Smid M, Alexandrov LB, et al. Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation. *Nat Commun*. 2019 Apr 15;10(1):1749.
19. Duran-Ferrer M, Clot G, Nadeu F, Beekman R, Baumann T, Nordlund J, et al. The proliferative history shapes the dna methylome of b-cell tumors and predicts clinical outcome. *HemaSphere*. 2019 Jun;3:376.
20. Candelaria M, Gallardo-Rincón D, Arce C, Cetina L, Aguilar-Ponce JL, Arrieta O, et al. A phase II study of epigenetic therapy with hydralazine and magnesium valproate to overcome chemotherapy resistance in refractory solid tumors. *Ann Oncol*. 2007 Sep;18(9):1529–38.
21. Oza AM, Matulonis UA, Alvarez Secord A, Nemunaitis J, Roman LD, Blagden SP, et al. A Randomized Phase II Trial of Epigenetic Priming with Guadecitabine and Carboplatin in Platinum-resistant, Recurrent Ovarian Cancer. *Clin Cancer Res*. 2020 Mar 1;26(5):1009–16.
22. Glasspool RM, Brown R, Gore ME, Rustin GJS, McNeish IA, Wilson RH, et al. A randomised, phase II trial of the DNA-hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in combination with carboplatin vs carboplatin alone in patients with recurrent, partially platinum-sensitive ovarian cancer. *Br J Cancer*. 2014 Apr 15;110(8):1923–9.
23. Wang YK, Bashashati A, Anglesio MS, Cochrane DR, Grewal DS, Ha G, et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat Genet*. 2017 Jun;49(6):856–65.
24. Eckert MA, Pan S, Hernandez KM, Loth RM, Andrade J, Volchenboum SL, et al. Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the fallopian tube. *Cancer Discov*. 2016 Oct 7;6(12):1342–51.
25. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep*. 2019 Jun 27;9(1):9354.
26. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet*. 2018 Apr 2;50(4):591–602.
27. Salhab A, Nordström K, Gasparoni G, Kattler K, Ebert P, Ramirez F, et al. A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. *Genome Biol*. 2018 Sep 28;19(1):150.
28. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D941–7.
29. Schroeder DI, Lott P, Korf I, LaSalle JM. Large-scale methylation domains mark a functional subset of neuronally expressed genes. *Genome Res*. 2011 Oct;21(10):1583–91.
30. Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res*. 2020 Jan 8;48(D1):D983–91.
31. Gong J, Wan H, Mei S, Ruan H, Zhang Z, Liu C, et al. Pancan-meQTL: a database to systematically evaluate the effects of genetic variants on methylation in human cancer. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D1066–72.

32. Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* 2016 Feb;26(2):256–62.
33. Aziz D, Etemadmoghadam D, Caldon CE, Au-Yeung G, Deng N, Hutchinson R, et al. 19q12 amplified and non-amplified subsets of high grade serous ovarian cancer with overexpression of cyclin E1 differ in their molecular drivers and clinical outcomes. *Gynecol Oncol.* 2018 Sep 9;151(2):327–36.
34. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014 May 15;30(10):1363–9.
35. Yao L, Shen H, Laird PW, Farnham PJ, Berman BP. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* 2015 May 21;16:105.
36. Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics.* 2019 Jun 1;35(11):1974–7.
37. Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol.* 2009 Apr;27(4):361–8.
38. Spainhour JC, Lim HS, Yi SV, Qiu P. Correlation patterns between DNA methylation and gene expression in the cancer genome atlas. *Cancer Inform.* 2019 Feb 11;18:1176935119828776.
39. Montavon C, Gloss BS, Warton K, Barton CA, Statham AL, Scurry JP, et al. Prognostic and diagnostic significance of DNA methylation patterns in high grade serous ovarian cancer. *Gynecol Oncol.* 2012 Mar;124(3):582–8.
40. Tomar T, Alkema NG, Schreuder L, Meersma GJ, de Meyer T, van Criekinge W, et al. Methylome analysis of extreme chemoresponsive patients identifies novel markers of platinum sensitivity in high-grade serous ovarian cancer. *BMC Med.* 2017 Jun 23;15(1):116.
41. Fang F, Cardenas H, Huang H, Jiang G, Perkins SM, Zhang C, et al. Genomic and Epigenomic Signatures in Ovarian Cancer Associated with Resensitization to Platinum Drugs. *Cancer Res.* 2018 Feb 1;78(3):631–44.
42. Lambrechts S, Smeets D, Moisse M, Braicu EI, Vanderstichele A, Zhao H, et al. Genetic heterogeneity after first-line chemotherapy in high-grade serous ovarian cancer. *Eur J Cancer.* 2016 Jan;53:51–64.
43. Coscia F, Lengyel E, Duraiswamy J, Ashcroft B, Bassani-Sternberg M, Wierer M, et al. Multi-level Proteomics Identifies CT45 as a Chemosensitivity Mediator and Immunotherapy Target in Ovarian Cancer. *Cell.* 2018 Sep 20;175(1):159-170.e16.
44. Liu Y, Xu J, Choi HH, Han C, Fang Y, Li Y, et al. Targeting 17q23 amplicon to overcome the resistance to anti-HER2 therapy in HER2+ breast cancer. *Nat Commun.* 2018 Nov 9;9(1):4718.
45. Reyes HD, Devor EJ, Warriar A, Newton AM, Mattson J, Wagner V, et al. Differential DNA methylation in high-grade serous ovarian cancer (HGSOC) is associated with tumor behavior. *Sci Rep.* 2019 Nov 29;9(1):17996.
46. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet.* 2016 Jul;48(7):758–67.
47. Esteves P, Dard L, Brillac A, Hubert C, Sarlak S, Rousseau B, et al. Nuclear control of lung cancer cells migration, invasion and bioenergetics by eukaryotic translation initiation factor 3F. *Oncogene.* 2020;39(3):617–36.
48. Xu J, Yong M, Li J, Dong X, Yu T, Fu X, et al. High level of CFTR expression is associated with tumor aggression and knockdown of CFTR suppresses proliferation of ovarian cancer in vitro and in vivo. *Oncol Rep.* 2015 May;33(5):2227–34.
49. Lawrenson K, Fonseca MAS, Liu AY, Segato Dezem F, Lee JM, Lin X, et al. A Study of High-Grade Serous Ovarian Cancer Origins Implicates the SOX18 Transcription Factor in Tumor Development. *Cell Rep.* 2019 Dec 10;29(11):3726-3735.e4.
50. Izar B, Tirosh I, Stover EH, Wakiro I, Cuoco MS, Alter I, et al. A single-cell landscape of high-grade serous ovarian cancer. *Nat Med.* 2020 Aug;26(8):1271–9.
51. Olbrecht S, Busschaert P, Qian J, Vanderstichele A, Loverix L, Van Gorp T, et al. High-grade serous tubo-ovarian cancer refined with single-cell RNA sequencing: specific cell subtypes influence survival and determine molecular subtype classification. *Genome Med.* 2021 Jul 9;13(1):111.
52. Prostate Cancer Biorepository Network. Prostate Cancer Biorepository Network SOP No. 006 RNA Extraction from Frozen Tissue [Internet]. 2016 [cited 2017 Apr 15]. Available from: https://prostatebiorepository.org/sites/default/files/2016%20PCBN%20SOPs/PCBN.SOP06.v2.0_RNA_Extraction_Frozen_Tissues.pdf
53. GitHub - huishenlab/biscuit: BISulfite-seq CUI Toolkit [Internet]. [cited 2021 Sep 29]. Available from: <https://github.com/huishenlab/biscuit>
54. Picard Tools - By Broad Institute [Internet]. [cited 2019 Jul 23]. Available from: <http://broadinstitute.github.io/picard/>

55. Babraham Bioinformatics - Trim Galore! [Internet]. [cited 2019 Jul 22]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
56. Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016 Oct 1;32(19):3047–8.
57. R: The R Project for Statistical Computing [Internet]. [cited 2021 Sep 29]. Available from: <https://www.r-project.org/>
58. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010 Nov 30;11:587.
59. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J*. 2016 Aug;8(1):289–317.
60. Korthauer K, Chakraborty S, Benjamini Y, Irizarry RA. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics*. 2019 Jul 1;20(3):367–83.
61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010 Mar 15;26(6):841–2.
62. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics*. 2017 Aug 1;33(15):2381–3.
63. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015 Jul 15;31(14):2382–3.
64. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D766–73.
65. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011 Aug 4;12:323.
66. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016 May;34(5):525–7.
67. Perteau M, Perteau GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015 Mar;33(3):290–5.
68. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017 Apr;14(4):417–9.
69. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. [version 2; peer review: 2 approved]. *F1000Res*. 2015 Jan 1;4:1521.
70. Jiménez-Sánchez A, Cast O, Miller ML. Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res*. 2019 Dec 15;79(24):6238–46.
71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
72. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005 Oct 25;102(43):15545–50.
73. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D590–5.
74. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012 May;16(5):284–7.

Figures

Figure 1

Clinical features of high grade serous ovarian cancer patients with recurrent disease: (A) Location of primary and recurrent tumors in women with and without germline *BRCA1/2* mutations; (B) Disease course for each patient. Filled circles indicate tumors that were profiled in this study, open circles tumors that were collected but not profiled; (C) Women carrying germline *BRCA1/2* mutations have an improved survival compared to non *BRCA1/2* mutation carriers; (D) Sample selection and QC process for inclusion in analysis.

Figure 2

High grade serous ovarian cancers show heterogeneous patterns of genome-wide methylation: (A) Both primary and recurrent tumors show heterogeneous patterns of methylation across the genome, with many tumors showing extensive hypomethylation on the X chromosome (CpG values are averaged across 10kB windows, minus ENCODE 'blacklist' regions); (B) Examples of two regions on chromosome 1q42.13 and 22q13.33, that show differentially methylated regions (boxed regions) from two comparisons - Primary vs Recurrent tumors (left) and *BRCA1/2* carrier vs *BRCA1/2* non-carrier (right)

Figure 3

Hypermethylation within partially methylated domains (PMDs) in high grade serous ovarian cancers is driven by soloWCGWs: (A) Illustration of PMD-masking strategy prior to calling differentially methylated regions. PMDs were identified as described in methods, and then those genomic regions were masked out of subsequent analyses. (B) Most PMDs detected across the cohort were unique to a single tumor, with only 2% of PMDs observed in more than 30 tumors; (C) Principal components (PC) analysis identifies a large proportion of the variance between tumors was due to methylation at soloWCGW sites within PMDs. Masking the genome for common PMDs and ovcaPMDs removed much of the variance; (D) Pairwise comparison for all possible tumor pairs. The strong correlation between PMD soloWCGW difference and pairwise Euclidean distance was lost after masking PMDs.

Figure 4

Methylation and transcription are largely preserved between primary and recurrent tumors from each patient, as shown by the expression of genes within partially methylated domains (PMDs): (A) Genes within PMDs shared by multiple tumor specimens are less expressed (left) but more variable in their expression (right) than genes outside of PMDs; (B) The vast majority of tumor suppressor genes in cancer and genes that form the ovarian cancer molecular subtypes defined by The Cancer Genome Atlas are located outside of PMDs; (C) Intra-patient pairwise Euclidean distances were significantly smaller than inter-patient distance or the intra-stage distance in both methylation (top) and gene expression (below) from paired RNA-Seq.

Figure 5

Significant differences in methylation and expression by *BRCA1/2* mutation carrier status: (A) 135 differentially methylated regions (DMRs) were identified in tumors from *BRCA1/2* carriers compared to non-carriers with a trend towards hypermethylation in tumors from non-carrier; 101 regions hypermethylated in tumors from *BRCA1/2* non-carriers compared to only 34 regions hypermethylated tumors from mutation carriers; (B) Principal components analysis using genome-wide CpG methylation level after PMD masking shows a trend towards differences in tumors based on *BRCA1/2* carrier status; (C) Principal components analysis of gene expression data shows a trend towards clustering of tumors based on *BRCA1/2* carrier status (D) Volcano plot of differentially expressed genes comparing tumors from *BRCA1/2* carriers vs non-carriers. Significantly up-regulated genes in tumors from *BRCA1/2* carriers are colored orange (Padj < 0.05), significantly down-regulated genes are colored purple; (E) KEGG gene set enrichment analysis for up-(orange) and down-(purple) differentially expressed genes in tumors from *BRCA1/2* carriers vs non-carriers; (F) Individual genes where methylation levels within hypermethylated regions in tumors from *BRCA1/2* non-carriers were correlated with gene expression (P-value < 0.005); *NAGLU*, *CDK2AP1*, *FGF18*.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1.xlsx](#)
- [AdditionalFile2.pptx](#)