A polygenic methylation prediction model associated with response to chemotherapy in epithelial ovarian cancer

Lanbo Zhao,[1,3] Sijia Ma,[1,3] Linconghua Wang,[2,3] Yiran Wang,[1] Xue Feng,[1] Dongxin Liang,[1] Lu Han,[1] Min Li,[2] and Qiling Li[1]

[1]Department of Obstetrics and Gynecology, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China; [2]Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, Hunan, China

To identify potential aberrantly differentially methylated genes (DMGs) correlated with chemotherapy response (CR) and establish a polygenic methylation prediction model of CR in epithelial ovarian cancer (EOC), we accessed 177 (47 chemo-sensitive and 130 chemo-resistant) samples corresponding to three DNA-methylation microarray datasets from the Gene Expression Omnibus and 306 (290 chemo-sensitive and 16 chemo-resistant) samples from The Cancer Genome Atlas (TCGA) database. DMGs associated with chemotherapy sensitivity and chemotherapy resistance were identified by several packages of R software. Pathway enrichment and protein-protein interaction (PPI) network analyses were constructed by Metascape software. The key genes containing mRNA expressions associated with methylation levels were validated from the expression dataset by the GEO2R platform. The determination of the prognostic significance of key genes was performed by the Kaplan-Meier plotter database. The key genes-based polygenic methylation prediction model was established by binary logistic regression. Among accessed 483 samples, 457 (182 hypermethylated and 275 hypomethylated) DMGs correlated with chemo resistance. Twenty-nine hub genes were identified and further validated. Three genes, anterior gradient 2 (AGR2), heat shock-related 70-kDa protein 2 (HSPA2), and acetyltransferase 2 (ACAT2), showed a significantly negative correlation between their methylation levels and mRNA expressions, which also corresponded to prognostic significance. A polygenic methylation prediction model (0.5253 cutoff value) was established and validated with 0.659 sensitivity and 0.911 specificity.

INTRODUCTION

Epithelial ovarian cancer (EOC) is the most common lethal gynecologic malignancy, and the most fatal cancer in the female reproductive system, with a 5-year survival rate of only 39%.[1–3] Late diagnosis and chemo resistance could account for its therapeutic failure and high mortality.[4] Currently, the standard treatment for patients with EOC is optimal cytoreductive surgery combined with platinum-based chemotherapy.[5,6] Nevertheless, although most patients initially respond to chemotherapy with complete or partial remission, approximately 80% of patients develop chemotherapy resistance, and up to 75% of patients eventually relapse within less than 2 years.[7–9] Acquired resistance to chemotherapeutic drugs is a major barrier to the treatment of EOC. It is significant to find biomarkers that can effectively and accurately predict responses to chemotherapy. However, until now, there are no available effective biomarkers to predict the effects of chemotherapy in patients with EOC.[10]

Epigenetic mechanisms have been proven to play a critical role in chemotherapy resistance of ovarian cancer.[11] Particularly, it has been demonstrated that abnormal DNA methylation—one of the most common and crucial epigenetic modifications—may affect the sensitivity of tumor cells to antitumor drugs by robustly modulating the expression of genes associated with chemotherapy response and serve as a potential biomarker to predict response to chemotherapeutic strategies.[12–14]

In the past few years, several studies have confirmed the relationship between gene hypermethylation and chemo resistance. The effect of gene methylation may depend on the location and level of gene methylation. Cacan[15] and Tomar et al.[16] separately discovered that gene RGS2 is hypermethylated in chemo-resistant ovarian cancer cells, and gene FAM83A and MYO18B are hypermethylated in patients with no response to chemotherapy. Tian et al.[17] first provided the evidence that due to hypermethylation of the upstream region, loss of expression of gene hMSH2 could play a vital role in the mechanism of platinum resistance in patients with EOC. Glasspool et al.[18] have illustrated that after using cytidine 5-azacytidine to successfully eliminate the methylation status of gene MLH1, silencing of MLH1

Table 1. Patient characteristics from TCGA and GEO database

Variables	Chemo-sensitive group	Chemo-resistant group	p value
Number	337	277	
Age (years)			
Mean	58.4	59	
Range	26−87	43−77	0.692
Stage			
I	1	218	
II	14	1	
III	252	25	0.675
IV	40	2	
Not available	30	31	
Grade			
G1	1	218	
G2	46	1	
G3	253	24	0.751
G4	0	1	
Not available	37	33	

was found to be closely related to platinum-based chemotherapy resistance in ovarian cancer cell lines. Other similar studies have also found that epigenetic silencing of SFRP5 may activate proto-oncogenes in the Wnt pathway, leading to progression of ovarian cancer and chemotherapy resistance.[19]

Although compared with hypermethylation, hypomethylation of ovarian cancer-related genes has been studied insufficiently, more attention has been paid in recent years. A study on the gene MAL has shown that MAL is highly expressed at the transcriptional level in platinum-resistant ovarian cancer cell lines, and the overexpression of MAL caused by promoter hypomethylation is correlated with poor prognosis in ovarian cancer.[20] Similarly, methylation-induced inactivation of the gene FANCF in ovarian cancer cell lines was followed by an increasing sensitivity to platinum.[21] Therefore, the detection of abnormal DNA methylation alterations is a promising tool for the prediction of chemotherapy response.

However, no DNA methylation gene was identified to have the high sensitivity and specificity to accurately represent the chemotherapy response of ovarian cancer. One reason may be that previous studies have mostly paid attention to one or a few candidate genes, which could not reflect the most significant feature of drug resistance. To successfully identify satisfactory biomarkers to predict responses to chemotherapy, big data analysis by comprehensive bioinformatics must be combined to select key genes and establish effective polygenic methylation prediction models.

The aim of our study is to explore specific DNA methylation genes as biomarkers and establish a polygenic methylation prediction model to effectively predict patients' responses to chemotherapy. Gene

Expression Omnibus (GEO)[22] datasets and The Cancer Genome Atlas (TCGA) database were combined to identify chemotherapy response-related differentially methylated genes (DMGs) of EOC, including hypermethylated and hypomethylated genes. Next, main hub genes were detected by investigating the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and protein-protein interaction (PPI) network analysis using Metascape software. Then, key genes were obtained through expression data analysis and prognostic validation. At last, through comprehensive analysis, an optimal key genes-based polygenic methylation prediction model was established to effectively predict individual responsiveness to chemotherapy in EOC.

RESULTS

Patient characteristics

The data from 483 samples with serous EOC were extracted from GEO and TCGA. The clinical characteristics of the discovery cohort are shown in Table 1. There were a total of 337 chemo-sensitive patients and 146 chemo-resistant patients. No significant difference was found in the age (p = 0.692), stage (p = 0.675), and grade (p = 0.751) between the two groups.

Datasets and sample selection

In accordance with the inclusion and exclusion criteria, 178 samples (corresponding to three datasets) from three GEO datasets and 306 samples of EOC from TCGA database were accessed. After removing 1 poor quality sample, 177 samples from GEO and 306 samples from TCGA were further analyzed.

Group assignments

In our study, a total of 483 samples, including 337 chemo-sensitive ovarian cancer samples and 146 chemo-resistant ovarian cancer samples, were analyzed, which were comprised of 177 samples from GEO datasets (47 chemo-sensitive samples and 130 chemo-resistant samples) and 306 samples from TCGA (290 chemo-sensitive samples and 16 chemo-resistant samples). We defined these two data clusters from GEO and TCGA database as follows (Figure S1):

Cluster 1: From GEO datasets (Illumina HumanMethylation450 BeadChip), mainly according to the citations from PubMed, samples were categorized into two groups: chemo-sensitive group (n = 47) and chemo-resistant group (n = 131), which was defined as cluster 1.

Cluster 2: From TCGA database, DNA methylation data from 626 patients were obtained (n = 10 using the Illumina HumanMethylation450 BeadChip and n = 616 using the Illumina HumanMethylation27 BeadChip). All patients underwent surgical treatment followed by platinum-based chemotherapy. After excluding patients whose clinical information of the response to the first therapy was not available, we divided the patients (n = 306) into two groups: a chemo-sensitive group (n = 290) and a chemo-resistant group (n = 16) according to the period after initial treatment. Due to the large number of differences between these two groups, the

direct comparison may not be statistically significant. Because it has been demonstrated that some same probes exist in both 27k and 450k BeadChip, those probes in the 450k BeadChip from the GEO datasets (specifically, from the chemo-resistant group in GEO) that correspond to the same probes in the 27k BeadChip were extracted and added to the chemo-resistant group in TCGA. There were a total of 130 samples from the GEO chemo-resistant group that were added to TCGA chemo-resistant group. Finally, cluster 2 was defined as follows: chemo-sensitive group (n = 290) cases (all from TCGA) and chemo-resistant group (n = 146) cases (16 from TCGA, 130 from GEO).

Data preprocessing

For cluster 1 from three GEO datasets, there were a total of 485,512 probes on the Illumina HumanMethylation450 BeadChip download. The minfi package of R software was used for data quality control with the following steps: one poor quality sample was excluded for p value cutoff (>0.05). Then, after normalization, 320 probes with a p value greater than 0.01 in at least 50% of samples, 11,346 probes on the X or Y chromosome, 23,343 probes containing SNPs, and 20,569 cross-reactive probes were excluded. Next, after using the sva package of R software and the combat function to batch effect, 429,934 probes (corresponding to 9,387 genes) were retained and further analyzed from the original 485,512 probes. The limma package of R software was chosen to perform the probes-level differential methylation analysis. Probes with adjusted p value <0.01 and beta value >0.2 or <−0.2 were used to identify significant differentially methylated probes (DMPs). At last, 38 probes (corresponding to 25 genes) were differentially methylated (Figure S1A).

For cluster 2 from the combination of GEO and TCGA, after quality control and probes affected by SNP and on X and Y chromosomes were removed, data were merged with the probes containing null values, and there were a total of 20,410 shared probes corresponding to 15,000 genes in the 450k and 27k BeadChip. Finally, with the use of the limma package of R software, 404 probes (corresponding to 432 genes) were differentially methylated (Figure S1B).

Detection of DMGs

The methylation differences between the chemo-sensitive group and the chemo-resistant group with the combination of cluster 1 and cluster 2 were mainly analyzed. After normalization and quality control, 429,934 probes (corresponding to 9,388 genes) in cluster 1 and 20,410 probes (corresponding to 15,000 genes) in cluster 2 were retained and further analyzed. With the use of the criteria of adjusted p value <0.01 and beta value >0.2 or <−0.2, 38 DMPs corresponding to 25 genes in cluster 1 and 404 DMPs corresponding to 432 genes in cluster 2 were identified. Among these, a total of 457 DMGs, 182 genes (40%), were significantly hypermethylated, and 275 genes (60%) were significantly hypomethylated in chemo-sensitive groups versus chemo-resistant groups.

With these DMGs, their regional distribution in the gene context, CpG-island (CGI) neighborhood, and chromosome was investigated,

respectively. First, the methylation level distribution of probes located in six gene-based regions (TSS1500, TSS200, 5′ UTR, first exon, gene body, and 3′ UTR) and six CGI-based regions (CGIs, south and north shores, south and north shelves, and OpenSea) was identified. Among the 182 hypermethylated genes, 23% genes' methylation sites were located in body and 5′ UTR, separately; 22% were located in the 1st exon; 19% were located in TSS1500; and 10% and 3% were located in TSS200 and 3′ UTR, respectively (Figure 1A). Referring to the six CGI-based regions, 47% of hypermethylated genes were located in OpenSea, and 22% were in island; the remaining was no less than 31% (Figure 1B). Among the 275 hypomethylated genes, 38% were located in TSS1500; 20% were in 5′ UTR; 16% were in body; and 15%, 10%, and 1% were in the 1st exon, TSS200, and 3′ UTR, respectively (Figure 1C). Referring to the six CGI-based regions, 46% were in OpenSea, 20% and 18% were in north shore and south shore, 12% were in island, and the remaining was no less than 5% (Figure 1D). Then, the distribution of probes located in the chromosome is shown in Figures 1E and 1F, respectively, among hypermethylated and hypomethylated genes. Several original missing data about chromosomal location were complemented by searching on the University of California, Santa Cruz (UCSC), Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly. There were significant differences found in the gene context, CGI neighborhood, and chromosome of regional distribution between hypermethylated and hypomethylated genes (p = 0.01, 0.000, and 0.009).

Pathway enrichment and PPI network analyses

To explore the biological functions of the 457 DMGs, GO, KEGG pathway enrichment, and PPI network analyses were performed by Metascape.

After calculating enrichment factors and accumulative hypergeometric p values, a subset of representative statistically enriched pathways (both GO and KEGG) were clustered and converted into a network layout based on Kappa-statistical similarities (Figure 2A). A specific pathway term was represented by a circle node, of which the number of input genes that fall into that term was in proportion to the size and of which cluster identity was represented by a specific color. Terms with a similarity score more than 0.3 were linked by an edge (the similar score was represented by the edge's thickness). The network was visualized by Cytoscape (v.3.1.2).

In GO Biological Processes, a total of 254 pathways were found. According to ranking these enrichment pathways by −log10(P), the top 10 pathways were the following: cellular defense response, positive regulation of apoptotic signaling pathway, response to extracellular stimulus, microglial cell-mediated cytotoxicity, cell-substrate adhesion, myeloid leukocyte activation, metal ion homeostasis, negative regulation of immune system process, neuron cellular homeostasis, and hindlimb morphogenesis. The heatmap of the top 10 GO pathways is shown in Figure 2B.

In KEGG, there were a total of five pathways enriched as follows: cytokine-cytokine receptor interaction, *Staphylococcus aureus* infection,
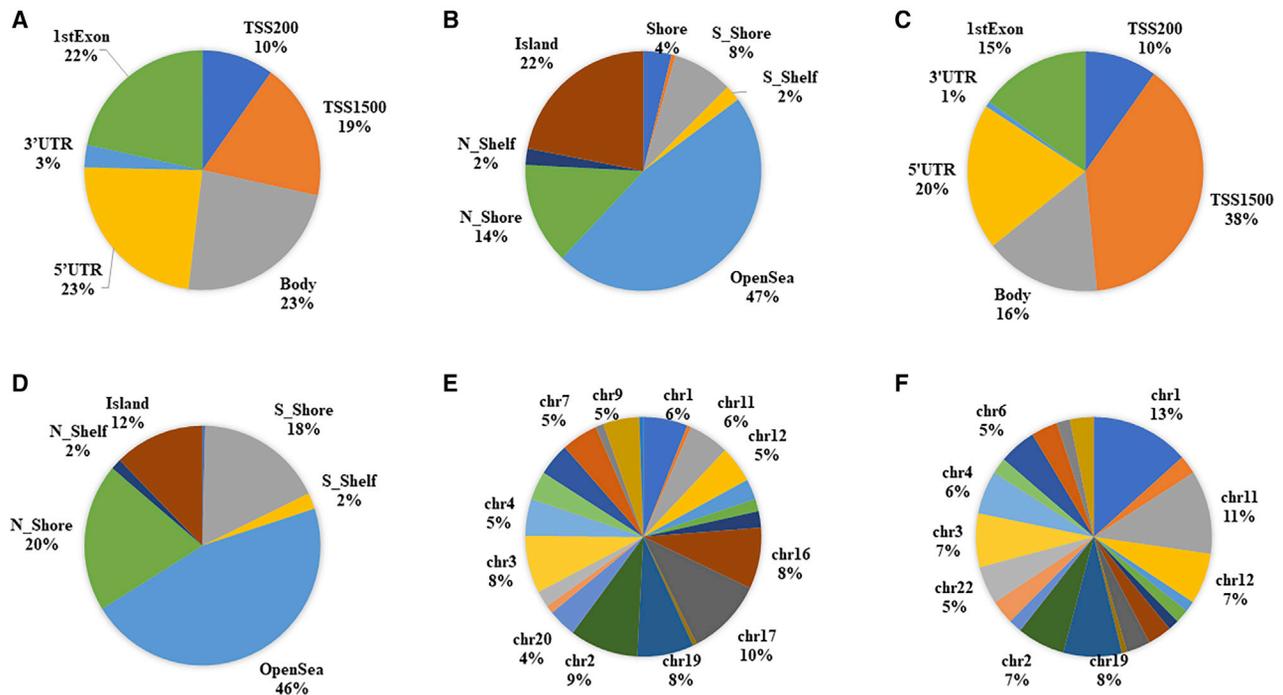
Figure 1. Regional distribution in the gene context, CpG-island neighborhood, and chromosome between hypermethylated and hypomethylated genes
(A) The proportion of hypermethylated and hypomethylated genes among all 457 differentially methylated genes. (B) Regional distribution in the gene context of 182 differentially hypermethylated genes. (C) Regional distribution in the CpG-island neighborhood of 182 differentially hypermethylated genes. (D) Regional distribution in the gene context of 275 differentially hypomethylated genes. (E) Regional distribution in the CpG-island neighborhood of 275 differentially hypomethylated genes. (F) Regional distribution in the chromosome of 182 differentially hypermethylated genes.

cell adhesion molecules (CAMs), calcium signaling pathway, and amphetamine addiction (p = 0.005, 0.004, 0.006, 0.007, and 0.008, respectively) (Table 2).

The MCODE algorithm in ClusterViz was then applied to identify neighborhoods where proteins were densely connected. Each MCODE network was assigned to a unique color (Figure 2C). MCODE 1, which had the most densely connected proteins, was composed of a nucleobase-containing small molecule metabolic process, ribonucleotide triphosphate metabolic process, and small molecule biosynthetic process. The biological interpretations about PPI network and MCODE components are in Table S1.

After comprehensively analyzing the gene data from pathway enrichment, and PPI network analysis combined with detailed gene studies manually searched in PubMed about the correlation between specific genes and cancer, 29 hub genes were selected for further validation. (Detailed selection process is shown in Figure S2.) Of these, we identified 10 genes whose beta value was the top 10 among 457 DMGs between the chemo-sensitive and chemo-resistant groups (DNMBP/NCAPH2/BLNK/TTYH1/ECEL1/NDST4/HAMP/anterior gradient 2 [AGR2]/TRAF3IP3/GBGT1); 8 genes belonging to the MCODE 1, which had the most densely connected proteins (heat shock-related 70-kDa protein 2 [HSPA2]/HIST1H2BK/PFKM/ARF4/acetyltransferase 2 [ACAT2]/PABPC3/ALDH1A3/EEF1B2); 5 genes belonging to

the *Staphylococcus aureus* infection pathway in KEGG (C2/CFB/HLA-DOB/ITGAL/ITGAM); and 6 genes belonging to the cellular defense response pathway in GO, which was the top 1 pathway in −log10(P) (ADORA2A/CD5L/CXCR2/LBP/LSP1/ZNF148). (Detailed gene information is shown in Table 3.)

The validation of hub gene expression levels
With the consideration of the relationship between abnormal DNA methylation and gene expression, we validated the selected hub genes with the expression data from GEO: GSM15372 (GEO expression profiling by array). According to the referring study, which provided the difference of gene expression between chemo-sensitive and chemo-resistant ovarian cancer cell lines,[21] ten samples in the study corresponding with the two groups above were defined, and the GEO2R platform was used to obtain the differentially expressed genes (DEGs) with the standard of log, fold change (FC) >1 or <−1, p <0.05.

Three keys genes, AGR2, HSPA2, and ACAT2, were validated to intuitively have a negative correlation between DNA methylation levels and RNA expressions among the sensitive and resistant groups. The histogram of RNA expressions from GEO2R and the boxplot of DNA methylation levels from the aforementioned bioinformatical analysis about AGR2, HSPA2, and ACAT2 are shown in Figures 3A, 3B, 3D, 3E, 3G, and 3H, respectively. In the chemo-resistant group, the RNA expression and DNA methylation
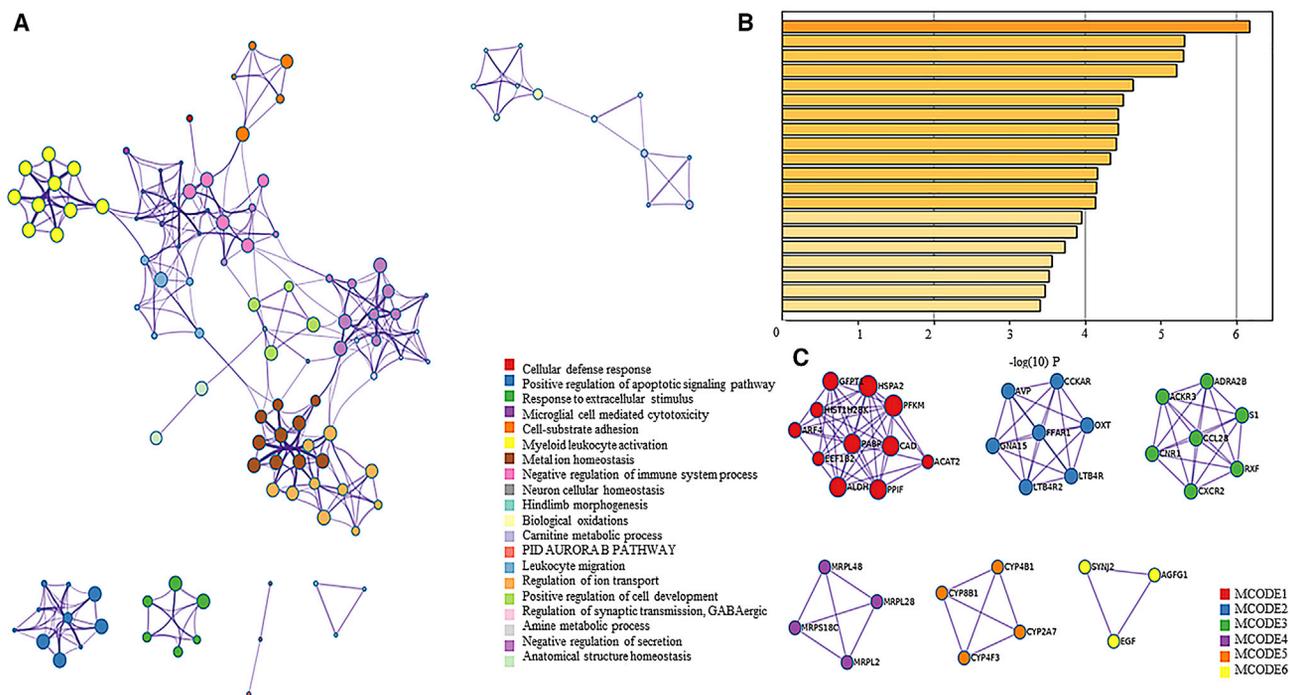
Figure 2. Enrichment pathway and PPI network analysis by Metascape

(A) Enriched ontology clusters colored by cluster ID. (B) Heatmap of top 10 GO pathways. (C) PPI MCODE components.

levels of gene AGR2 were lower and higher compared with the chemo-sensitive group, which was in accordance with the worse prognosis. Genes HSPA2 and ACAT2 showed a similar tendency, in which the RNA expression and DNA methylation levels of genes HSPA2 and ACAT2 were higher and lower in the chemo-resistant group with worse prognosis.

The prognostic significance of validated key genes

To estimate the prognostic significance of abnormally expressed AGR2, HSPA2, and ACAT2, the survival time (including progression-free survival [PFS] and post-progression survival [PPS]) and gene-expression levels were acquired from the Kaplan-Meier plotter website (Figures 3C, 3F, and 3I, respectively). The analysis results showed that the high expression level of AGR2 and lower level of HSPA2 and ACAT2 were related to longer PFS and PPS, which was consistent with the resistant status.

Establishment and validation of the polygenic methylation prediction model

Based on methylation data from these three validated key genes from 473 samples from GEO and TCGA, a polygenic methylation prediction model was established to effectively predict patients' responses to chemotherapy. Because of the independent variables of the three genes' DNA methylation levels, which were continuous variables, and the dependent variable of response to chemotherapy (i.e., chemo sensitive or chemo resistant), which is a binary or dichotomous variable, the binary logistic regression (LR) model was appropriate and

was chosen to establish the model. To validate the accuracy and efficiency of the model, we randomly divided the samples into training set (n = 146) and validation set (n = 337). Next, the forward: LR method in binary LR was selected to help screen for independent variables that had a significant impact on the dependent variable. Three models and corresponding detailed information, including correct rate, area under the curve (AUC), 95% confidence interval (CI), cutoff value, sensitivity, and specificity among the training set and validation set, are shown in Table 4. The receiver operating characteristic[23] curves of the three models with training set and validation set are shown in Figures 4A and 4B, respectively. The third model, which included all three key genes as independent variables, had the largest AUC of 0.84 and the highest correction prediction rate of 82.1%, which suggested its high accuracy in prediction performance. The cutoff value of the third model was 0.5253 with 0.659 sensitivity and 0.911 specificity.

$$\text{Model 1}: \log it(P) = 15.594 * AGR2 - 3.387$$

$$\text{Model 2}: \log it(P) = 12.619 * AGR2 - 6.023 * ACAT2 - 1.713$$

$$\text{Model 3}: \log it(P) = 15.011 * AGR2 + 5.983 * HPSA2 - 5.797 \\ * ACAT2 - 4.219$$

Therefore, it was obvious that the third model was the most appropriate polygenic methylation prediction model to predict responses to chemotherapy with high accuracy and efficiency.

Table 2. KEGG enrichment pathway

GO	Description	log(P)	Hits
hsa04060	cytokine-cytokine receptor interaction	−2.34116	TNFRSF17|EGF| IL3|CXCR2|OSM| PRLR|CX3CL1|GDF5| CCL26|IL20RA| CCL28|ACKR3
hsa05150	*Staphylococcus aureus* infection	−2.43601	CFB|C2|HLA-DOB|ITGAL|ITGAM
hsa04514	cell adhesion molecules (CAMs)	−2.25605	CD6|HLA-DOB|ITGAL| ITGAM|SPN|CLDN16|CLDN15|NLGN2
hsa04020	calcium signaling pathway	−2.16711	ADORA2A|ATP2B2|CCKAR| GNA15|GNAS|P2RX1|PPP3R2|PPIF|LTB4R2
hsa05031	amphetamine addiction	−2.07771	GNAS|PPP1CC|PPP3R2|CREB5|GRIN3A

DISCUSSION

It has been suggested that the alterations of DNA methylation-induced expression of various drug response-related genes and pathways may take part in the development of chemotherapy resistance in EOC.[17] In the present study, we used a series of bioinformatical methods to screen the most likely chemotherapy response-related DMGs in EOC. With the data combination between GEO and TCGA, 457 genes (corresponding to 182 hypermethylated genes and 275 hypomethylated genes) were identified. Metascape software was chosen to analyze the pathway enrichment (GO and KEGG) and PPI network among the genes to find dense connections among these genes. After comprehensive analysis, 29 hub genes were identified to be further validated. By the combination of expression data from GEO and prognostic information from a survival curve, three key genes (AGR2, HSPA2, and ACAT2) were validated to have a negative correlation between DNA methylation levels and RNA expressions, and their RNA expressions among the sensitive and resistant groups had corresponding prognoses.

Based on these three genes' methylation levels, three prediction models were established and validated by binary LR with the forward: LR method. AUC, sensitivity, and specificity were calculated to verify the accuracy of models, respectively. The most optimal prediction was confirmed by the highest AUC.

AGR2 belongs to a family of chaperone-like proteins, namely, protein disulfide isomerase (PDI), which are micro-environmentally regulatory proteins that can catalyze the formation, reduction, or isomerization of disulfide bonds in their network. These enzymatic reactions promote the maturation of proteins into bioactive conformations in the endoplasmic reticulum (ER).[24,25] Recently, it has been demonstrated that downregulation of AGR2 was associated with progression and chemotherapy resistance in ovarian cancer.[23] HSPA2 is a member of the heat shock protein 70 family,[26] and ACAT2 belongs to lipid metabolism enzymes.[27] Although, at present, there has been no research available illustrating the relation among gene HSPA2, ACAT2, and drug resistance in ovarian cancer, it has been shown that HSPA2 is associated with tumor progression, and overexpression of HSPA2 was proven to correlate with tumor angiogenesis and poor prognosis in pancreatic carcinoma.[27] Furthermore, ACAT2 plays a

pivotal role in the prognosis of clear cell renal cell carcinoma.[28] These studies indirectly validated our bioinformatical results.

Finally, an optimal polygenic methylation prediction model was established in our study. There are few previous studies focusing on establishing a prediction model to predict chemotherapy responses in ovarian cancer. In 2016, Gonzalez Bosquet et al.[29] were the first to construct prediction models to predict chemotherapy response by applying multiple different modeling methods. But, compared with our results, although Gonzalez Bosquet et al.[29] applied nine statistical methods to perform the model, the highest AUC among those nine models was only 0.73, which was lower than any model in our study. Apart from using only AUC and 95% CI to measure the performance of prediction models, we also calculated sensitivity and specificity between the training set and validation set. Through detecting methylation levels of specific sites in these three genes and obtaining the prediction value, these models could not only be used to predict patients' response to chemotherapy before treatment, which could guide the clinician to choose effective drug therapy, but also could be used during the treatment as effect biomarkers to monitor patients' responses and reflect the transition of acquired drug resistance in a timely way.

Admittedly, limited by the current data information of tumor databases, which have few available samples concentrating specifically on the drug sensitivity and resistance of ovarian cancer, we only have 483 samples analyzed and only refer to the level of gene methylation and RNA expression with different samples. To validate and improve our models in the future, more sequencing data are urgently needed. Since EOC is a kind of very heterogeneity cancers, to discover the whole landscape of chemotherapy resistance, the use of DNA methylation and RNA expression only is largely insufficient; multiomics sequencing, including whole genome sequencing, transcriptome sequencing, epigenetic sequencing, proteome sequencing, and metabolome sequencing, must be combined and detected on the same sample to eliminate the heterogeneity and establish multiomics prediction models.

Our study has established an optimal polygenic methylation prediction model, based on three key genes—AGR2, HSPA2, and

Table 3. The characteristics of 29 hub genes

P_to_T gene	P_to_T adj. p val	P_to_T logFC	Gene context	CpG-island neighborhood	Chromosome
C2	1.48E−38	−0.214045444	3′ UTR	OpenSea	chr6
CFB	1.48E−38	−0.214045444	TSS1500	OpenSea	chr6
HLA-DOB	2.19E−17	−0.219949192	body	OpenSea	chr6
ITGAL	2.54E−31	0.20677909	1st exon	N_shore	chr16
ITGAM	1.21E−24	0.215777262	body	OpenSea	chr16
DNMBP	6.83E−138	0.445191907	5′ UTR	OpenSea	chr10
NCAPH2	6.10E−165	−0.440596603	1st exon; 5′ UTR	island	chr22
BLNK	9.59E−67	−0.365733256	1st exon	OpenSea	chr10
TTYH1	1.60E−80	−0.352957924	TSS1500	N_shore	chr19
ECEL1	5.65E−60	0.344504731	TSS1500	island	chr2
NDST4	1.35E−54	−0.344206777	TSS1500	OpenSea	chr4
HAMP	9.92E−43	0.33090596	1st exon; 5′ UTR	S_shelf	chr19
AGR2	1.97E−64	−0.326537378	5′ UTR; 1st exon	OpenSea	chr7
TRAF3IP3	2.29E−93	0.325696324	TSS200	OpenSea	chr1
GBGT1	1.80E−34	0.324487248	5′ UTR	island	chr9
HSPA2	3.91E−51	0.291132114	1st exon	island	chr14
HIST1H2BK	9.99E−21	0.232543124	3′ UTR	N_shore	chr6
PFKM	2.85E−22	−0.21989798	5′ UTR; 1st exon; body	OpenSea	chr12
ARF4	2.05E−112	−0.215217182	TSS1500	S_shore	chr3
ACAT2	5.02E−31	0.233758973	TSS1500	island	chr6
PABPC3	1.43E−14	0.200421216	TSS1500	island	chr13
ALDH1A3	1.13E−51	0.299991937	body	island	chr15
EEF1B2	6.45E−91	0.219721961	body	S_shore	chr2
ADORA2A	4.77E−32	−0.212979809	5′ UTR	OpenSea	chr22
CD5L	1.54E−24	−0.236469294	TSS200	OpenSea	chr1
CXCR2	1.49E−27	0.204648037	TSS200; 5′ UTR	OpenSea	chr2
LBP	6.70E−62	0.204763054	body	OpenSea	chr20
LSP1	7.87E−38	0.234401803	5′ UTR;1st exon	OpenSea	chr11
ZNF148	1.50E−149	−0.298332627	5′ UTR	island	chr3

adj. p val, adjusted p value; N, north; S, south.

ACAT2—to predict patients' response to chemotherapy and help clinicians choose effective drug therapy before and during treatment.

MATERIALS AND METHODS
Source of microarray data
GEO methylation datasets and TCGA genomic data, including DNA methylation and clinical information, which contain the periods from initial therapy to recurrence, were downloaded, normalized, formatted, and organized for the analysis, according to the precepts of the data-sharing agreements from GEO and TCGA. The methylation data from GEO and TCGA are open ended and publicly available, mainly comprised of Illumina Infinium Human DNA Methylation 450k and 27k arrays separately (Illumina, San Diego, CA, USA).

Search strategy and selection criteria
All ovarian cancer datasets and samples with clinical and methylation information from GEO and TCGA were screened, filtered, and selected by hand. The inclusion criteria were as follows: (1) samples were ovarian cancer tissues from patients with EOC; (2) data were comprised of DNA methylation arrays from Illumina HumanMethylation450 or 27 BeadChip; and (3) complete follow-up information was included to clearly estimate patients' responses to chemotherapy. The exclusion criteria were as follows: (1) samples were from cell lines or patient-derived xenograft (PDX); (2) data were not from methylation profiling or not from Illumina Human-Methylation450 or 27 BeadChip; (3) data were non-primary downloaded data; and (4) samples were primary resistant, refractory, or using demethylation drugs.
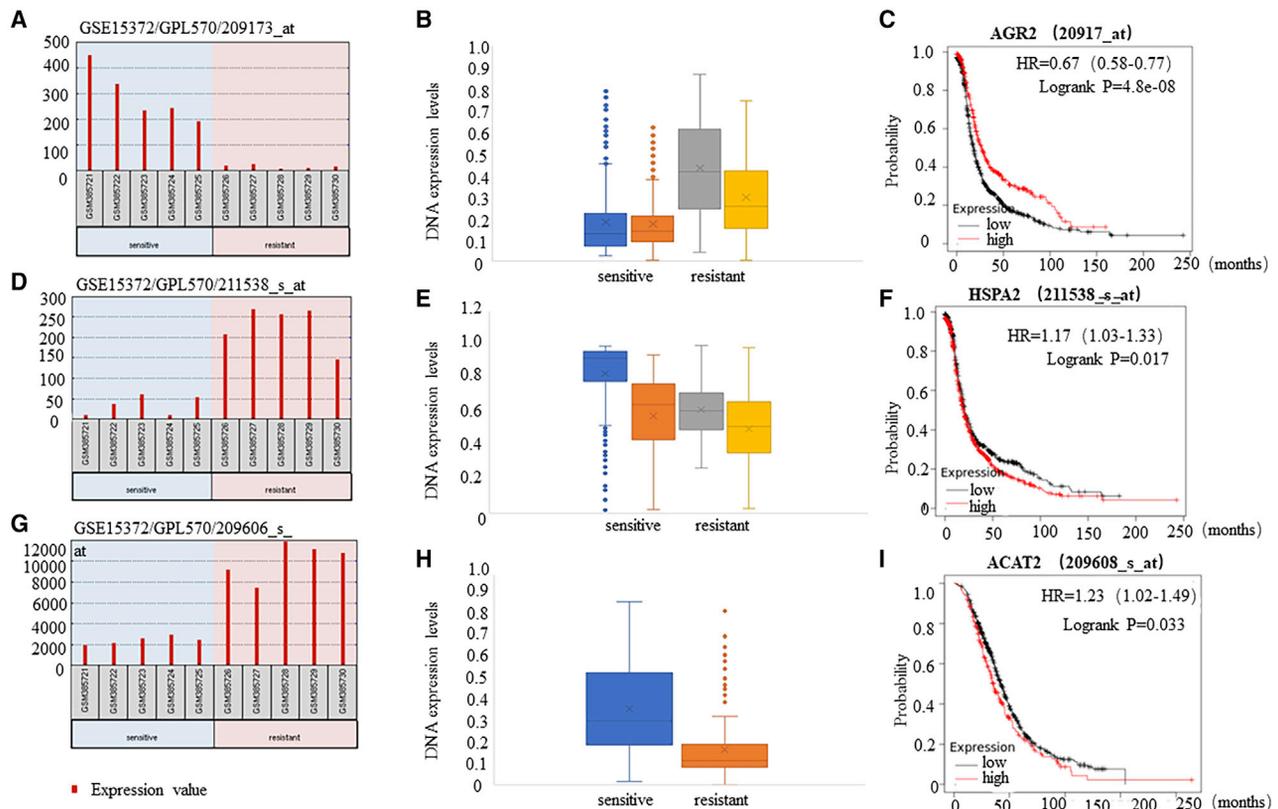
Figure 3. The DNA methylation levels, RNA expressions, and survival curves among three key genes

(A) RNA expression levels between the sensitive and resistant group about gene AGR2. (B) DNA methylation levels between the sensitive and resistant group about gene AGR2. (C) Survival curves based on gene expression level and survival time (PFS) about gene AGR2. (D) RNA expression levels between the sensitive and resistant group about gene HSPA2. (E) DNA methylation levels between the sensitive and resistant group about gene HSPA2. (F) Survival curves based on gene expression level and survival time (PFS) about gene HSPA2. (G) RNA expression levels between the sensitive and resistant group about gene ACAT2. (H) DNA methylation levels between the sensitive and resistant group about gene ACAT2. (I) Survival curves based on gene expression level and survival time (PPS) about gene ACAT2.

Group assignments

According to the clinical information and follow-up data downloaded from GEO and TCGA databases, samples were divided into two groups: a chemo-sensitive group and a chemo-resistant group. The chemo-sensitive group was defined as those with no evidence of disease progression within 6 months after the completion of front-line treatment.[22] The chemo-resistant group was those who had a recurrence within 6 months after treatment completion, having previously demonstrated sensitivity to earlier lines of chemotherapy.[22,30–32]

Data processing

DNA methylation data with beta values were downloaded from the GEO and TCGA data portals, extracted, loaded, and normalized to extract significantly DMPs. The differential DNA methylation of genes was calculated based on beta values.[33]

Pathway enrichment and PPI network analyses

To further characterize the molecular characteristics of DMGs, GO Biological Processes and KEGG Pathway enrichment and PPI network analysis were performed by Metascape software: http://metascape.org/gp/index.html#/main/step1,[34] and ClusterViz.[35] Hub genes were identified by comprehensively considering the results from beta values, pathway enrichment, and PPI network analysis.

Validation of the expression of hub genes and correlation between methylation levels and mRNA expressions

RNA expression levels from the gene expression dataset (GEO: GSE15372) were extracted. The GEO2R platform was used to detect DEGs between chemo-sensitive and chemo-resistant samples among the hub genes. The key genes were identified by the negative correlation between methylation levels and mRNA expressions.

Prognostic analysis of key genes

The PFS or PPS curves of each key gene were drawn using the Kaplan-Meier plotter platform: http://kmplot.com/analysis/index.php.[36] The hazard ratio (HR), 95% CI, and log rank p value were evaluated.

Table 4. The detailed information of three prediction models between training set and validation set

Binary logistic regression (LR)	Training set							Validation set			
	Cox & Snell R square	Correct rate	AUC	95% CI	Cutoff	Sensitivity	Specificity	AUC	95% CI	Sensitivity	Specificity
Forward_LR−step 1	0.2	73.80%	0.79	0.715−0.865	0.2241	0.818	0.663	0.838	0.790−0.887	0.922	0.441
Forward_LR−step 2	0.262	79.30%	0.82	0.730−0.901	0.3121	0.818	0.743	0.853	0.805−0.901	0.912	0.53
Forward_LR−step 3	0.291	82.10%	0.84	0.762−0.918	0.5253	0.659	0.911	0.852	0.804−0.900	0.873	0.667

Establishment and validation of polygenic methylation prediction model

The polygenic methylation prediction model was established to effectively predict an EOC patient's response to chemotherapy. The dependent variable that was classified into sensitive (negative) and resistant (positive) groups was a binary or dichotomous variable. Since the binary regression models are built to predict the function of binary- or dichotomous-dependent variables as predictive variables, they are applicable methods for predicting the likelihood of a positive or negative diagnosis. Sensitivity, specificity, AUC, and 95% CI were calculated to measure the performance of the prediction model. The binary LR model is as follows:

$$\log it(P) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

The p value is the probability that y is equal to 1.

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}$$

Statistical analysis

After normalization, all methylation data were analyzed by R 3.1.2 software: https://www.r-project.org/.[37] Quality control and DMPs were identified by Bioconductor minfi and limma packages of R software, respectively. In accordance with the platform annotation file, the DMPs were annotated into corresponding DMGs. The Benjamini-Hochberg false-discovery rate (FDR) method of adjusted p value of each gene was calculated. The DMGs were screened out as adjusted p value less than 0.01 and beta value either greater than 0.2 (termed differentially hypermethylated gene) or less than −0.2 (termed differentially hypomethylated gene). The parameters for the DEGs were set with |log2FC| >1 and adjusted p <0.05. The statistical analysis was performed using SPSS v.22.0 software (SPSS, Chicago, IL, USA). The comparisons between two groups were performed with the Student's t test, chi-square test, and Mann-Whitney rank test. p <0.05 was considered statistically significant.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omto.2021.02.012.
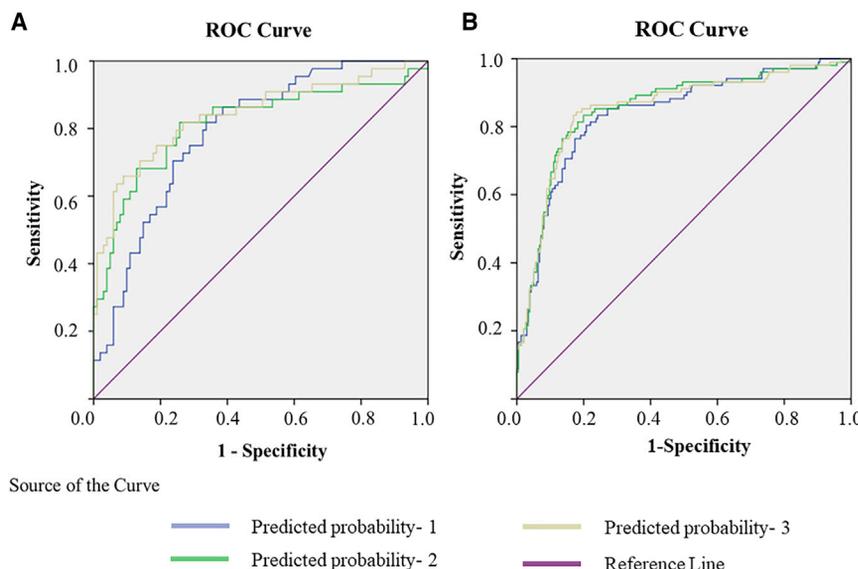
Figure 4. Receiver operating characteristic (ROC) curves of three models with training set and validation set

(A) Three ROC curves of three models with training set. (B) Three ROC curves of three models with validation set.

AUTHOR CONTRIBUTIONS

Q.L. and M.L. designed and configured this study. L.Z., S.M., and L.W. performed the majority of the experiments. Y.W. and X.F. prepared the figures. S.M. wrote the manuscript. Q.L. and L.Z. revised the manuscript. D.L. and L.H. helped to collect the data.

DECLARATION OF INTERESTS

REFERENCES

1. Taylor, K.N., and Eskander, R.N. (2018). PARP Inhibitors in Epithelial Ovarian Cancer. Recent Patents Anticancer Drug Discov. 13, 145–158.

2. Siegel, R.L., Miller, K.D., and Jemal, A. (2019). Cancer statistics, 2019. CA Cancer J. Clin. 69, 7–34.

3. Chen, H.J., Huang, R.L., Liew, P.L., Su, P.H., Chen, L.Y., Weng, Y.C., Chang, C.C., Wang, Y.C., Chan, M.W.Y., and Lai, H.C. (2018). GATA3 as a master regulator and therapeutic target in ovarian high-grade serous carcinoma stem cells. Int. J. Cancer 143, 3106–3119.

4. Lengyel, E. (2010). Ovarian cancer development and metastasis. Am. J. Pathol. 177, 1053–1064.

5. de Leon, M., Cardenas, H., Vieth, E., Emerson, R., Segar, M., Liu, Y., Nephew, K., and Matei, D. (2016). Transmembrane protein 88 (TMEM88) promoter hypomethylation is associated with platinum resistance in ovarian cancer. Gynecol. Oncol. 142, 539–547.

6. Kwon, J.S., McGahan, C., Dehaeck, U., Santos, J., Swenerton, K., and Carey, M.S. (2014). The significance of combination chemotherapy in epithelial ovarian cancer. Int. J. Gynecol. Cancer 24, 226–232.

7. Alberts, D.S. (1999). Treatment of refractory and recurrent ovarian cancer. Semin. Oncol. 26 (1, Suppl 1), 8–14.

8. Martin, L.P., and Schilder, R.J. (2009). Management of recurrent ovarian carcinoma: current status and future directions. Semin. Oncol. 36, 112–125.

9. Public Health Agency of Canada; Statistics Canada; Canadian Cancer Society; provincial/territorial cancer registries (2019). Release notice - Canadian Cancer Statistics 2019. Health Promot. Chronic Dis. Prev. Can. 39, 255.

10. Mase, S., Shinjo, K., Totani, H., Katsushima, K., Arakawa, A., Takahashi, S., Lai, H.C., Lin, R.I., Chan, M.W.Y., Sugiura-Ogasawara, M., and Kondo, Y. (2019). ZNF671 DNA methylation as a molecular predictor for the early recurrence of serous ovarian cancer. Cancer Sci. 110, 1105–1116.

11. Borley, J., and Brown, R. (2015). Epigenetic mechanisms and therapeutic targets of chemotherapy resistance in epithelial ovarian cancer. Ann. Med. 47, 359–369.

12. Natanzon, Y., Goode, E.L., and Cunningham, J.M. (2018). Epigenetics in ovarian cancer. Semin. Cancer Biol. 51, 160–169.

13. Wolffe, A.P., and Matzke, M.A. (1999). Epigenetics: regulation through repression. Science 286, 481–486.

14. Chang, X., Monitto, C.L., Demokan, S., Kim, M.S., Chang, S.S., Zhong, X., Califano, J.A., and Sidransky, D. (2010). Identification of hypermethylated genes associated with cisplatin resistance in human cancers. Cancer Res. 70, 2870–2879.

15. Cacan, E. (2017). Epigenetic regulation of RGS2 (Regulator of G-protein signaling 2) in chemoresistant ovarian cancer cells. J. Chemother. 29, 173–178.

16. Tomar, T., Alkema, N.G., Schreuder, L., Meersma, G.J., de Meyer, T., van Criekinge, W., Klip, H.G., Fiegl, H., van Nieuwenhuysen, E., Vergote, I., et al. (2017). Methylome analysis of extreme chemoresponsive patients identifies novel markers of platinum sensitivity in high-grade serous ovarian cancer. BMC Med. 15, 116.

17. Tian, H., Yan, L., Xiao-Fei, L., Hai-Yan, S., Juan, C., and Shan, K. (2019). Hypermethylation of mismatch repair gene hMSH2 associates with platinum-resistant disease in epithelial ovarian cancer. Clin. Epigenetics 11, 153.

18. Glasspool, R.M., Brown, R., Gore, M.E., Rustin, G.J., McNeish, I.A., Wilson, R.H., Pledge, S., Paul, J., Mackean, M., Hall, G.D., et al.; Scottish Gynaecological Trials Group (2014). 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 110, 1923–1929.

19. Niskakoski, A., Kaur, S., Staff, S., Renkonen-Sinisalo, L., Lassus, H., Järvinen, H.J., Mecklin, J.P., Bützow, R., and Peltomäki, P. (2014). Epigenetic analysis of sporadic and Lynch-associated ovarian cancers reveals histology-specific patterns of DNA methylation. Epigenetics 9, 1577–1587.

20. Pruszynski, M., Koumarianou, E., Vaidyanathan, G., Revets, H., Devoogdt, N., Lahoutte, T., Lyerly, H.K., and Zalutsky, M.R. (2014). Improved tumor targeting of anti-HER2 nanobody through N-succinimidyl 4-guanidinomethyl-3-iodobenzoate radiolabeling. J. Nucl. Med. 55, 650–656.

21. Li, M., Balch, C., Montgomery, J.S., Jeong, M., Chung, J.H., Yan, P., Huang, T.H.M., Kim, S., and Nephew, K.P. (2009). Integrated analysis of DNA methylation and gene expression reveals specific signaling pathways associated with platinum resistance in ovarian cancer. BMC Med. Genomics 2, 34.

22. Patch, A.M., Christie, E.L., Etemadmoghadam, D., Garsed, D.W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P.J., et al.; Australian Ovarian Cancer Study Group (2015). Whole-genome characterization of chemoresistant ovarian cancer. Nature 521, 489–494.

23. Alves, M.R., E Melo, N.C., Barros-Filho, M.C., do Amaral, N.S., Silva, F.I.B., Baiocchi Neto, G., Soares, F.A., de Brot Andrade, L., and Rocha, R.M. (2018). Downregulation of AGR2, p21, and cyclin D and alterations in p53 function were associated with tumor progression and chemotherapy resistance in epithelial ovarian carcinoma. Cancer Med. 7, 3188–3199.

24. Bergström, J.H., Berg, K.A., Rodríguez-Piñeiro, A.M., Stecher, B., Johansson, M.E., and Hansson, G.C. (2014). AGR2, an endoplasmic reticulum protein, is secreted into the gastrointestinal mucus. PLoS ONE 9, e104186.

25. Sevier, C.S., and Kaiser, C.A. (2002). Formation and transfer of disulphide bonds in living cells. Nat. Rev. Mol. Cell Biol. 3, 836–847.

26. Bonnycastle, L.L., Yu, C.E., Hunt, C.R., Trask, B.J., Clancy, K.P., Weber, J.L., Patterson, D., and Schellenberg, G.D. (1994). Cloning, sequencing, and mapping of the human chromosome 14 heat shock protein gene (HSPA2). Genomics 23, 85–93.

27. Zhai, L.L., Xie, Q., Zhou, C.H., Huang, D.W., Tang, Z.G., and Ju, T.F. (2017). Overexpressed HSPA2 correlates with tumor angiogenesis and unfavorable prognosis in pancreatic carcinoma. Pancreatology 17, 457–463.

28. Zhao, Z., Lu, J., Han, L., Wang, X., Man, Q., and Liu, S. (2016). Prognostic significance of two lipid metabolism enzymes, HADHA and ACAT2, in clear cell renal cell carcinoma. Tumour Biol. 37, 8121–8130.

29. Bosquet, Jesus, Gonzalez, Newtson, Andreea, M, Chung, Rebecca, K, Thiel, Kristina, W, Ginader, Timothy, and Goodheart, Michael, J (2016). Prediction of chemoresponse in serous ovarian cancer.

30. Friedlander, M.L., Stockler, M.R., Butow, P., King, M.T., McAlpine, J., Tinker, A., and Ledermann, J.A. (2013). Clinical trials of palliative chemotherapy in platinum-resistant or -refractory ovarian cancer: time to think differently? J. Clin. Oncol. 31, 2362.

31. Therasse, P., Arbuck, S.G., Eisenhauer, E.A., Wanders, J., Kaplan, R.S., Rubinstein, L., Verweij, J., Van Glabbeke, M., van Oosterom, A.T., Christian, M.C., and Gwyther, S.G. (2000). New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J. Natl. Cancer Inst. 92, 205–216.

32. Friedlander, M., Butow, P., Stockler, M., Gainford, C., Martyn, J., Oza, A., Donovan, H.S., Miller, B., and King, M. (2009). Symptom control in patients with recurrent ovarian cancer: measuring the benefit of palliative chemotherapy in women with platinum refractory/resistant ovarian cancer. Int. J. Gynecol. Cancer *19* (*Suppl 2*), S44–S48.

33. Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

34. Metascape http://metascape.org/gp/index.html#/main/step1.

35. Wang, J., Zhong, J., Chen, G., Li, M., Wu, F.X., and Pan, Y. (2015). ClusterViz: A Cytoscape APP for Cluster Analysis of Biological Network. IEEE/ACM Trans. Comput. Biol. Bioinformatics *12*, 815–822.

36. KM-plotter (2009–2021). Kaplan-Meier Plotter, http://kmplot.com/analysis/index.php.

37. R Development Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing), www.r-project.org/.