

# The Establishment of an Ovarian Cancer Omentum Metastasis Related Prognostic Model by Integrated Analysis of ScRNA-Seq and Bulk RNA-Seq.

**Dongni Zhang**

Guang'anmen Hospital

**Wenping Lu** (✉ [lu\\_wenping@sina.com](mailto:lu_wenping@sina.com))

Guang'anmen Hospital

**Shasha Cui**

Guang'anmen Hospital

**Heting Mei**

Guang'anmen Hospital

**Xiaoqing Wu**

Guang'anmen Hospital

**Zhili Zhuo**

Guang'anmen Hospital

---

## Research Article

**Keywords:** ovarian cancer, scRNA-seq, omentum metastasis, 6-OMAGs, risk, prognosis

**Posted Date:** March 2nd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1393066/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Objective

Ovarian cancer is a gynecological malignant tumor with the highest mortality, it preferentially metastasizes to omental tissue, leading to intestinal obstruction and death. scRNA-Seq is a powerful technique to analyze tumor heterogeneity. Analyzing omentum metastasis of ovarian cancer at the single cell level may be more conducive to analyzing and understanding omentum metastasis and prognosis of ovarian cancer at the cellular function and genetic levels.

## Methods

The scRNA-seq data of omentum metastasis site were acquired from the GEO (Gene Expression Omnibus) database, the data were clustered by Seurat package and annotated by SingleR package. Cell differentiation trajectories were reconstructed through monocle package, and obtained the marker genes of every branch. The microarray data of ovarian cancer were extracted from GEO and were clustered into omentum metastasis associated subpopulations according to the marker genes of single cell differentiation trajectories. The tumor microenvironment and immune infiltration differences between subpopulations were analyzed by estimate package and CIBERSORT. After that, based on the survival information and the intersection of bulk RNA-seq data of TCGA (The Cancer Genome Atlas) and marker genes of GEO subpopulations, univariate Cox and Lasso regression were performed to further establish an omentum metastasis-associated genes (OMAGs) signature, the signature was then tested by GEO data. Finally, the clinicopathological characteristics of TCGA-OV and GEO were screened by univariate and multivariate Cox regression analysis, in order to draw the nomogram.

## Results

The scRNA-seq data (GSE147082) were clustered into 18 cell clusters, and annotated into 14 cell types. Reconstruction of differentiation trajectories divided the cells into 5 branches, a total of 833 cell trajectories related characteristic genes were obtained. Two microarray datasets (GSE132342 and GSE135820) were integrated to gain the expression data of 510 genes in 7846 patients, which were subtyped into 3 subpopulations by the cell trajectories related characteristic genes. K-M survival analysis showed that the prognosis of cluster2 was the worst and cluste1 was the best,  $p < 0.001$ . The tumor microenvironment showed that the ESTIMATE Score, Stromal Score in cluster 2 is significantly higher than that in the other two clusters, while Immune Score, Tumor Purity were significantly lower,  $p < 0.001$ ; The immune infiltration analysis showed differences in the fraction of 10 immune cells among the 3 subpopulations,  $p < 0.05$ . Immune checkpoint analysis showed that there were significant differences in the expression levels of 6 genes in the 3 subpopulations, which were significantly correlated with prognosis. Subsequently, the scRNA-seq characteristic genes of 5 branches, GEO characteristic genes of 3 subpopulations were intersected with the bulk RNA-seq of 379 patients in TCGA-OV, and 11 candidate

genes of OMAGs (omentum metastasis-associated genes) were obtained by Lasso regression. The composition of OMAGs was further refined by Cox multivariate regression analysis, and established a 6-OMAGs signature ( $\text{RiskScore} = 6.669 \times \text{VCAN} + 0.173 \times \text{FN1} - 0.128 \times \text{IGKC} + 0.258 \times \text{RPL21} + 2.133 \times \text{UCL1} + 0.457 \times \text{RARRES1}$ ). Taking TCGA data as training set and geo data as test set, ROC curves were drawn to verify its sensitivity. Ultimately, 5 clinicopathological characteristics (age, tumor status, initial treatment outcome, stage, grade) were screened by univariate and multivariate Cox regression analysis, and the nomogram was drawn.

## Conclusions

Finally, we constructed a clinical prognostic model related to omentum metastasis of ovarian cancer, which was composed of 6-OMAGs and 5 clinicopathological features.

## 1. Introduction

Ovarian cancer is the gynecological malignant tumor with the highest mortality[1]. At present, no effective method for early detection of ovarian cancer has been found. 70% of clinical ovarian cancer is advanced at the time of diagnosis, while after surgery and chemotherapy, 70% of patients will still have metastasis within 2–3 years[2]. Ovarian cancer metastasis has distinct characteristics: it preferentially metastasizes to omental adipose tissue, leading to intestinal obstruction and death[3]. However, the mechanism of this tropism is still remained unclear. Finding the cause of this transfer and preventing it has been a crucial problem that scholars hope to solve for many years.

The omentum is a special adipose tissue in the peritoneal cavity, which is composed of adipocytes and stromal blood vessels, including preadipocytes, fibroblasts, vascular endothelial cells and a variety of immune cells can promote various immune responses, including inflammation, tolerance and fibrosis, thereby promoting peritoneal immunity[4]. Studies have found that after the initial cytoreductive surgery, there are still residual cancer cells latent. The "latent cancer cells" may flow into the omentum through the production of substances such as inflammatory factors, resulting in the change of omental microenvironment: epithelial mesenchymal transformation (EMT), angiogenesis, immune infiltration, inflammation, etc.[5–8], drive the formation of niche before metastasis and contribute to successful transmission, that is, pre metastatic microenvironment (PMN). When tumor cells are planted in the omentum, they can support tumor growth through immune and metabolic mechanisms[7]. However, the preference for omentum preferential metastasis of ovarian cancer is still unclear.

Single cell RNA sequencing (scRNA-seq) uses optimized next-generation sequencing technology to define the global gene expression profile of single cells, which is helpful to isolate the previously hidden heterogeneity in the cell population, especially for the study of tumor heterogeneity[9]. In recent studies, Hu et al.[10] identified 6 subtypes of fallopian tube epithelium cells in normal human fallopian tube tissues by scRNA-seq, revealed intra-tumoral heterogeneity in serous ovarian cancer and defined SOC subtypes that correlate with patient prognosis; Olalekan et al.[11] using scRNA-seq from ovarian tumors

resected from omental metastases, revealed immune cell types and subsets with possible roles in the management of disease; Hornburg et al.[12] used scRNA-seq to describe the cell spacing of ovarian cancer microenvironment and shed light on how tumor, immune, and stromal cells interact and shape T cell infiltration.

In the present study, scRNA-seq data of ovarian cancer omentum metastasis site was used to explore significant significantly different genes related to prognosis in different cell clusters, combined with GEO microarray data and TCGA-OV bulk RNA-seq data, a risk model of 6-OMAGs (omentum metastasis-associated genes) signature was constructed. Integrated with the clinicopathological information of GEO and TCGA-OV, an ovarian cancer omentum metastasis related prognostic model of risk score and 5 clinical parameters was established. We believe our results will improve our understanding of molecular mechanism and characteristics of ovarian cancer omentum metastasis and its connection with patient survival.

## 2. Material And Methods

### 2.1 Data acquisition

The scRNA-seq data was downloaded from GEO (Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov/geo/>) database, accession number GSE147082, including 9,885 cells 16041 genes isolated from the omental metastatic site of 6 ovarian cancer patients. The bulk RNA-seq data of 379 ovarian cancer patients 56461 genes in TCGA-OV were obtained from TCGA (The Cancer Genome Atlas, <https://portal.gdc.cancer.gov/>) database. The microarray data were acquired from GEO with accession number GSE132342 and GSE135820, which contained 3769 and 4077 patients of 510 genes, respectively .

### 2.2 scRNA-seq data processing and analysis

By using the R package Seurat[13] (Versio:4.0.6), the scRNA-seq data was converted into a Seurat object by the function `CreateSeuratObject()`, while the `PercentageFeatureSet()` function is used to calculate the percentage of mitochondrial genes, unqualified data was rejected by the criteria: 1) cells less than 50 genes were detected; 2) genes identified in less than 3 cells; 3) mitochondrial DNA reads for more than 5%. The data was then normalized with the method "LogNormalize". The top 1500 genes with large coefficient of variation between cells were extracted by the function `FindVariableFeatures()`. Subsequently, PCA (principal component analysis) was performed, in order to evaluate the most significant principal components for cell cluster analysis. Next, cell cluster classification was assessed by t-NSE (t-Distributed Stochastic Neighbor Embedding), after that, the maker genes of each clusters were screened by the function `FindAllMarkers()` with the cutoff values of  $|\log_2 \text{fold change(FC)}| \geq 1$ , the expression ratio of cell population  $\leq 0.25$  and  $\text{adjust Pvalue} \leq 0.05$ . Afterwards, singleR[14] (Versio:1.8.0) pakage was utilized for cell clusters annotation, while the reference was loaded from `celldex` package, `HumanPrimaryCellAtlasData()`. Cell differentiation trajectories was reconstructed through `monocle`[15]

(Version:2.22.0) package, cell trajectory difference between each cluster was also analyzed, followed by the differential expression analysis and function enrichment, GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes).

## **2.3 GEO microarray data processing and analysis**

Both datasets GSE132342 and GSE135820 were already log2 transformed, since they used the same platform (NanoString nCounter Human CodeSet OTTA2014), the two gene expression count data were merged and the batch effecter was removed then standardized by using Limma[16] (Version: 3.50.0) package. Based on the results of cell cluster classification and cell trajectory analysis of scRNA-seq previously, the 7846 patients in total were divided into different groups via ConsensusClusterPlus[17] (Version: 1.58.0) package, determining the number and membership of possible clusters in the data set provides quantitative evidence, then Kaplan Meier survival curve and histogram of clinicopathological characteristics as well as the up and down-regulated gene in branches were showed to find the differences between groups.

## **2.4 Tumor microenvironment analysis**

In tumor microenvironment, immune cells and stromal cells are two main types of tumor cells, by using the expression data, estimate algorithm can predict the score of stromal cells and immune cells, thereby predict the content of these two cells; Finally, the tumor purity in each tumor sample can be calculated[18]. The ESTIMATE Score, Immune Score, Stromal Score, Tumor Purity of different groups of GEO patients were calculated respectively by using the estimate (<https://bioinformatics.mdanderson.org/estimate/rpackage.html>) R package, and visualized by violin plot.

## **2.5 Immune infiltration analysis**

CIBERSORT is an analytical tool developed by Newman et al.[19] to provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data. With R package “e1071” (Version: 1.7-9) loaded as a precondition, the CIBERSORT was used to estimate the different abundance of 22 immune cells among different groups of GEO microarray count data, based on the grouping results of scRNA-seq data analysis above, in order to further analyze the clinical significance of immune cells infiltrated by omentum metastasis ovarian cancer. K-M curve of high and low level of B cells naive, Macrophages M2, T cells CD4 memory activated, T cells CD4 naive separately, to show differences in survival.

## **2.6 Immune checkpoint analysis**

Several important immune checkpoints related with cancer were analyzed by Limma[16] (Version: 3.50.0) package. Differentially expressed immune checkpoint-related genes significantly associated with overall survival (OS) in clustered GEO patients were determined by K-M survival analysis.

## **2.7 Construction of OMAGs Molecular Risk Model**

The marker genes of different branches of scRNA seq were crossed by the genes in TCGA bulk seq and GEO microarray data, the intersection genes of the 3 datasets were taken as the candidate genes for OMAGs. The expression of these candidate genes in TCGA-OV were combined with the survival data for Lasso regression analysis, to obtain the genes related to prognosis. The glmnet[20] (Version: 4.1-3) and survival (Version: 3.2–13) R package was used to perform lasso Cox regression for analysis and 10-fold cross-validation to narrow the range of OMAGs (omentum metastasis-associated genes) candidate genes. It is a biased estimation for processing data with multicollinearity that can realise the selection of variables while estimating parameters to better solve the problem of multicollinearity in regression analysis. The OMAGs candidates were then screened again by Cox multivariate regression analysis in order to determine the ultimate OMAGs risk model. The calculation of the OMAGs signature was as follows: Risk score =  $\sum (\beta_i \times \text{Exp}_i)$ , where  $\beta_i$  represented the coefficient of gene  $i$ , standing for the weight of gene  $i$ , and  $\text{Exp}_i$  represented the expression level of gene  $i$ . Then the TCGA-OV dataset was set as the training set while GEO dataset as the test set, the effects of high and low risk scores on survival and the prognostic value of OMAGs risk model were evaluated.

## 2.8 Construction of a nomogram model in accordance with the OMAGs Risk Model

The clinicopathological features and risk score of TCGA-OV cohort were analyzed by univariate Cox regression and multivariate Cox regression using survival (Version: 3.2–13) R package, with  $p < 0.05$  as the criterion, and the HR and regression coefficient for each prognostic features were calculated. Finally, the rms (Version: 6.2-0) R package was used to construct a nomogram to predict OS of ovarian cancer patients, which incorporated these clinicopathological features and OMAGs. The nomogram model was a prognostic statistical model made using simple graphs according to previous studies.

## 3. Results

### 3.1 scRNA-seq data processing and analysis

A total of 9885 high-quality single cell samples were acquired from GSE147082. The quality control chart was shown in Fig. 2A, which illustrates the range of scRNA quantity, the sequencing counts, and the proportion of mitochondrial sequencing counts of each cell. 1500 variable genes were isolated from single cell samples, and the top 10 genes with the most significant difference were shown in Fig. 2B. The differences among 6 patients' samples were showed by PCA chart (Fig. 2C). PCA was carried out based on highly variable genes (Fig. 2D,E), these single cell samples were divided into 20 different components, based on the standard deviation using ElbowPlot function (Fig. 2F), JackStraw function was used to resample the test and the p values of each PCs were calculated to find statistically significant components (Fig. 2G). Since there was no obvious elbow point, and the p values were all  $< 0.01$ , we selected 14 PCs for downstream analysis. the t-NSE algorithm was used for non-linear dimension reduction and successfully clustered the single-cell samples into 18 clusters (Fig. 2H). Afterwards, the 18

cell clusters were annotated by singleR package label.fine, and a total of 14 cells types were recognized (Fig. 2I).

## **3.2 Reconstruction of differentiation trajectories of ovarian cancer omentum metastasis site**

The fate decisions and differentiation trajectories of ovarian cancer omentum metastasis were reconstructed with the Monocle package. The pseudotime, cluster, cell type and state estimation analysis of single cells (Fig. 3A-D) were performed based on the highly variable marker genes at the screening criteria of  $|\log_2 \text{fold change(FC)}| \geq 1$  and  $\text{adjust p value} < 0.05$ . Among them, branch 1 has 320 characteristic genes, branch 2 has 11 genes, branch 3 has 462 genes, branch 4 has 282 genes and branch 5 has 289 genes. By combining these genes, a total of 833 differentiation trajectories of ovarian cancer omentum metastasis characteristic genes were obtained.

## **3.3 Identification of OMAGs in ovarian cancer omentum metastasis using scRNA-seq data and GEO data**

The 2 datasets GSE132342 and GSE135820 were merged, with the removal of batch effect and normalization, a total of 77846 patients' microarray data of 510 genes and clinicopathology information were obtained. The GEO microarray data were clustered according to the 833 differentiation trajectories of ovarian cancer omentum metastasis characteristic genes in the previous analysis. The calculation was repeated 50 times using the ConsensusClusterPlus package, setting  $k_{\text{max}} = 9$ . According to the output results, the K value was determined as 3. Geo patients were divided into 3 subpopulations, and the characteristic genes of the 3 clusters were obtained. K-M curve was shown for survival analysis (Figure. 4A), with  $p < 0.01$ , indicating that there was significant survival difference among the 3 clusters of patients. It can be seen that the prognosis of the patients in cluster 2 is the worst, which is significantly different from the patients in cluster 1 and 3,  $P < 0.01$ , while the patients in cluster 1 and 3 have certain differences after 8 years of long-term survival. The histogram showed the diversity clinical information of the patients in 3 clusters. It can be seen from the histogram of patient clinical feature analysis that the prognosis of category 2 is the worst in which the tumor stage of the patients was later (Figure. 4B), the diagnosis age was higher (Figure. 4C), the proportion of omentum at the site of surgical resection was significantly higher than that in the other two clusters (Figure. 4D), and the sample source omentum was also higher than that in the other two clusters (Figure. 4E). Figure. 4F shows the expression difference of representative characteristic genes in the 5 branches of GEO patients, in which there were no significantly up-regulated or down-regulated genes in branch 2.

## **3.4 Analysis of tumor microenvironment and immune cell infiltration using GEO data**

The tumor microenvironment analysis was carried out on the gene expression data of 3 clusters of GEO patients using the estimate R package. The results are shown in Fig. 5A. Showed that among cluster 2 with the worst prognosis, the estimate score and matrix score were significantly the highest, and the immune score and tumor purity score were the lowest. Then the CIBERSORT package was used for immune infiltration analysis, and only the cells with more accurate classification ( $P < 0.05$ ) were selected for comparison. The proportion of immune cells in 22 cells in each group is shown in Fig. 5B, and the proportion difference is shown in Fig. 5C. It can be seen that the content of Plasma cells, T cells CD8, T cells CD4 naïve, T cells follicular helper, NK cells activated, Monocytes, Macrophages M1, Mast cells activated, Eosinophils, Neutrophils cells in 3 clusters were significantly different. Then, the K-M survival curves of 4 immune cells whose proportion is significantly related to survival and prognosis are drawn respectively. The Fig. 5D showed that the higher proportion of T cells CD4 Naïve and B cells Naïve has a better prognosis, while the higher proportion of Macrophages M2 and T cells CD4 memory activated has a worse prognosis.

### 3.5 Immune checkpoint analysis

A list of 38 immune checkpoint genes was obtained from the literature search, which intersected with GEO microarray data, and the expression difference between 3 clusters of the intersecting genes were analyzed. 6 immune checkpoints with significant difference were screened. Their expression levels in the 3 clusters were shown in Fig. 6A. Geo patients were divided into high expression and low expression groups according to the expression levels of the 6 differential immune checkpoints, K-M curves were drawn for survival analysis (Fig. 6B-G). It can be seen that high expression of CD8A, CD274, CTLA4, JAK2 and PDCD1 is associated with better prognosis, while high expression of IDH8 has lower early survival rate and better expression, and the survival rate decreases about 9 years, which is lower than that of low expression group.

### 3.6 Screening the OMAGs and construct a molecular prognosis prediction model

The genes in TCGA-OV cohort, GEO cohort, and the 833 characteristic genes of 5 branches obtained from scRNA-seq analysis above were intersected, and a total of 76 candidate genes of OMAGs were obtained. 11 candidate OMAGs genes were screened by Lasso regression (Fig. 7A-B), which were "VCAN", "TIMP3", "FN1", "COLA2", "IGKC", "GSTP1", "RPL21", "UCHL1", "BNIP3", "RARRES1" and "IGF2". Then, multivariate Cox regression analysis was carried out to obtain 6 prognosis related genes "VCAN", "FN1", "IGKC", "RPL21", "UCHL1", "RARRES1". The coef value and P value of each gene were shown in Table 1. According to the expression level of 6-OMAGs, a molecular risk prognostic evaluation model was constructed, TCGA data is used as the training set (Fig. 7C), and geo data is used as the verification set (Fig. 7D). The relationship between the expression of 6-OMAGs and survival was verified respectively. It can be seen that the survival rate with high-risk score was lower and the survival rate with low-risk score was higher in both the training set and the test set. Then, the ROC curve was drawn and the AUC area was calculated. It

can be seen that the sensitivity in the training set was 0.58 in 3 years, 0.684 in 5 years and 0.801 in 10 years (Fig. 7E), it increases with time, and the sensitivity in the test set is lower than that in the verification set. It may be because one is high-throughput sequencing data and the other is microarray data, which belong to different cohorts, therefore, there is a certain batch effect.

Table 1  
The coef value and pvalue of the 6-OMAGs

ID	coef	HR	HR.95L	HR.95H	pvalue
VCAN	6.668590835	787.2854077	9.741963423	63623.55167	0.00292216
FN1	0.173180557	1.189080783	1.053293958	1.342372751	0.005122717
IGKC	-0.127767185	0.880058246	0.813871806	0.95162716	0.001360535
RPL21	0.258144129	1.294525384	1.073943409	1.560413663	0.006760239
UCHL1	2.132910074	8.439390361	1.43812874	49.52498873	0.018157007
RARRES1	0.456691178	1.578841229	1.140789919	2.185099627	0.005879973

### 3.7 Construction of a nomogram model in accordance with the 6-OMAGs signature

Univariate Cox analysis on the clinicopathology characteristics of TCGA-OV patients was performed. The results were shown in Fig. 8A, in which age, Tumor status (with or without tumor), the initial treatment outcome and risk score have obvious statistical significance. Subsequently, multivariate Cox analysis was carried out on these clinical characteristics. The results were shown in Fig. 8B, which proved that age, tumor status, initial treatment outcome and risk score are the influencing factors of prognosis. It may be that most patients are stage III-IV and grade 3–4, so the grade of tumor stage's p value > 0.05, but considering that it is a clinically recognized factor clearly related to tumor prognosis, it is still included in nomogram mapping. After that, we analyzed the expression levels of 6-OMAGs in 18 clusters of scRNA-seq of ovarian cancer omentum metastases. The results were shown in Fig. 8C-D. VCAN and FN1 had the highest expression in cluster1, corresponding to epithelial cells; The expression of IGKC was the highest in cluster 12, corresponding to B cells; RPL21 expressed in all clusters, but the highest expression in cluster11, corresponding to B cell Naïve; UCHL1 was highly expressed in cluster 4, but very low in other clusters, cluster 4 corresponds to epithelial cells; RARRES1 was significantly higher in clusters 4 and 6 than in other clusters, corresponding to epithelial cells and fibroblasts. Finally, we mapped nomograms related to the prognosis of ovarian cancer according to risk scores and clinicopathology characteristics (Fig. 8E). A horizontal line was drawn to determine the point of each variable, and the total points of each ovarian cancer patient, which were distributed between 0 and 100, was calculated by summing the points of all variables. Finally, a vertical line between the total points axis and each prognostic axis was utilized to evaluate the 1-, 3-, and 5-year OS of ovarian cancer patients.

## 4. Discussion

In this study, we combined scRNA-seq data of 6 omentum metastasis sites from 6 ovarian cancer patients, GEO microarray data of 7846 patients, TCGA-OV bulk RNA-seq data of 379 patients and clinical information to construct a prognosis prediction model of ovarian cancer composed of 6-OMAGs signature and 5 clinicopathological features.

As we can know from the previous researches, VCAN (versican) is a chondroitin sulfate proteoglycan[21], while proteoglycans possibly participate in tumor progression and metastasis, whereas they are ubiquitous constituent of extracellular matrix and cell surface[22]. Clinical trials as well as in vivo and in vitro experiments have verified that VCAN is related to the formation of tumor microenvironment and partly epigenetically regulated, can form a prognostic model together with other genes, which is completely consistent with our results[23–27].

FN1 (fibronectin), together with transforming growth factor beta-induced protein and tenascin C had been proved some related to transcoelomic spreading via the peritoneal fluid or malignant ascites[28], moreover, networks built with FN1 and CDKN1A as the center were significantly relevant with epithelial-to-mesenchymal transition, could be hallmarks to taxanes resistance in ovarian cancer cells[29], also used as a marker and target in tumor prognosis and treatment[30–32].

IGKC (immunoglobulin kappa C), expressed in plasma cells, had been identified as one of the top genes of a prognostic B cell-metagenes in breast cancer, correlated to favorable prognosis and response to chemotherapy[33], study showed plasma cell infiltration in epithelial ovarian cancer has a significant impact on tumor progression and prognosis[34]. High expression of IGKC is bound up with good outcome[35], its coef value is negative therefore negatively correlated with risk score in our model, which is consistent with previous studies.

As we know, ribosomal proteins (RPs) are involved in the cellular process of translation, and in recent years, RPs' dysfunction was found in tumors, like mutation, expression levels changes and correlation with differentiation[36–38]. RPL21 (ribosomal protein gene 21) has been found to be associated with the proliferation of pancreatic cancer cells[39] and may be used as a potential marker for cervical intraepithelial neoplasia 1[40], but there are hardly any reports in ovarian cancer, our findings may provide a reference for finding new mechanisms and therapeutic targets for omentum metastasis of ovarian cancer.

UCHL1 (ubiquitin carboxyl-terminal esterase L1) is an oncogene encoding a deubiquitinating enzyme which adjusts the balance between mTOR complexes[41], and plays a significant role in ubiquitin system as well as different cellular processes[42]. It has been widely studied and confirmed related to tumor progression, such as prostate cancer, lymphoma, lung cancer, breast cancer, neuroblastoma, and so on[42–45]. In ovarian cancer, UCHL1 was found may be one of the markers related to early stages of high-grade serous carcinoma, immunogenicity[46, 47], and cisplatin resistance[48, 49]. Our results also suggest its value in omentum metastasis and prognosis of ovarian cancer.

RARRES1 (retinoic acid receptor responder 1) is one of the common methylated loci in several cancers and believed to be a putative tumor suppressor gene[50, 51], promoter hypermethylation and loss of RARRES1 seemed to facilitate to cancer progression[52]. However, its specific role in omentum metastasis of ovarian cancer remains to be clarified, which needs further research.

Although there are various models related to the prognosis of ovarian cancer, few focus on omentum metastasis. Our model suggests the potential mechanism of omentum metastasis of ovarian cancer to a certain extent. There are still many deficiencies in our research, such as lack of first-hand data and no experimental verification, so more clinical data are needed to test it to improve its accuracy in the follow-up research.

## Declarations

### Acknowledgements:

None.

### Authors' contributions:

Dongni Zhang wrote the main manuscript text and run bioinformatics analysis, Wenping Lu guided and modified the manuscript, Shasha Cui retrieved and downloaded data, Xiaoqing Wu, Heting Mei and Zhili Zhuo prepared figures1-8. All authors reviewed the manuscript.

### Funding:

Beijing Municipal Natural Science Foundation, 7212192.

### Availability of data and materials:

Please contact the corresponding author Wenping Lu (lu\_wenping@sina.com).

### Competing interests:

None.

## References

1. Menon U, Gentry-Maharaj A, Burnell M et al. Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* 2021; 397: 2182-2193.
2. Torre LA, Trabert B, DeSantis CE et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin* 2018; 68: 284-296.

3. Pradeep S, Kim SW, Wu SY et al. Hematogenous metastasis of ovarian cancer: rethinking mode of spread. *Cancer Cell* 2014; 26: 77-91.
4. Lee W, Ko SY, Mohamed MS et al. Neutrophils facilitate ovarian cancer premetastatic niche formation in the omentum. *J Exp Med* 2019; 216: 176-194.
5. Girolimetti G, De Iaco P, Procaccini M et al. Mitochondrial DNA sequencing demonstrates clonality of peritoneal implants of borderline ovarian tumors. *Mol Cancer* 2017; 16: 47.
6. Curtis M, Kenny HA, Ashcroft B et al. Fibroblasts Mobilize Tumor Cell Glycogen to Promote Proliferation and Metastasis. *Cell Metab* 2019; 29: 141-155 e149.
7. Etzerodt A, Moulin M, Doktor TK et al. Tissue-resident macrophages in omentum promote metastatic spread of ovarian cancer. *J Exp Med* 2020; 217.
8. Zhang AW, McPherson A, Milne K et al. Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell* 2018; 173: 1755-1769 e1722.
9. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol* 2018; 14: 479-492.
10. Hu Z, Artibani M, Alsaadi A et al. The Repertoire of Serous Ovarian Cancer Non-genetic Heterogeneity Revealed by Single-Cell Sequencing of Normal Fallopian Tube Epithelial Cells. *Cancer Cell* 2020; 37: 226-242 e227.
11. Olalekan S, Xie B, Back R et al. Characterizing the tumor microenvironment of metastatic ovarian cancer by single-cell transcriptomics. *Cell Rep* 2021; 35: 109165.
12. Hornburg M, Desbois M, Lu S et al. Single-cell dissection of cellular components and interactions shaping the tumor immune phenotypes in ovarian cancer. *Cancer Cell* 2021; 39: 928-944 e926.
13. Hao Y, Hao S, Andersen-Nissen E et al. Integrated analysis of multimodal single-cell data. *Cell* 2021; 184: 3573-3587 e3529.
14. Aran D, Looney AP, Liu L et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019; 20: 163-172.
15. Qiu X, Mao Q, Tang Y et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017; 14: 979-982.
16. Ritchie ME, Phipson B, Wu D et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
17. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; 26: 1572-1573.
18. Yoshihara K, Shahmoradgoli M, Martinez E et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013; 4: 2612.
19. Newman AM, Steen CB, Liu CL et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019; 37: 773-782.
20. Tibshirani R, Bien J, Friedman J et al. Strong rules for discarding predictors in lasso-type problems. *J R Stat Soc Series B Stat Methodol* 2012; 74: 245-266.

21. Voutilainen K, Anttila M, Sillanpaa S et al. Versican in epithelial ovarian cancer: relation to hyaluronan, clinicopathologic factors and prognosis. *Int J Cancer* 2003; 107: 359-364.
22. Kusumoto T, Kodama J, Seki N et al. Clinical significance of syndecan-1 and versican expression in human epithelial ovarian cancer. *Oncol Rep* 2010; 23: 917-925.
23. Yamaguchi K, Mandai M, Oura T et al. Identification of an ovarian clear cell carcinoma gene signature that reflects inherent disease biology and the carcinogenic processes. *Oncogene* 2010; 29: 1741-1752.
24. Cheon DJ, Tong Y, Sim MS et al. A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clin Cancer Res* 2014; 20: 711-723.
25. Ghosh S, Albitar L, LeBaron R et al. Up-regulation of stromal versican expression in advanced stage serous ovarian cancer. *Gynecol Oncol* 2010; 119: 114-120.
26. Yeung TL, Leung CS, Wong KK et al. TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res* 2013; 73: 5016-5028.
27. Salem M, O'Brien JA, Bernaudo S et al. miR-590-3p Promotes Ovarian Cancer Growth and Metastasis via a Novel FOXA2-Versican Pathway. *Cancer Res* 2018; 78: 4175-4190.
28. Steitz AM, Steffes A, Finkernagel F et al. Tumor-associated macrophages promote ovarian cancer cell migration by secreting transforming growth factor beta induced (TGFBI) and tenascin C. *Cell Death Dis* 2020; 11: 249.
29. Duran GE, Wang YC, Moisan F et al. Decreased levels of baseline and drug-induced tubulin polymerisation are hallmarks of resistance to taxanes in ovarian cancer cells and are associated with epithelial-to-mesenchymal transition. *Br J Cancer* 2017; 116: 1318-1328.
30. Berg HF, Ju Z, Myrvold M et al. Development of prediction models for lymph node metastasis in endometrioid endometrial carcinoma. *Br J Cancer* 2020; 122: 1014-1022.
31. Gardi NL, Deshpande TU, Kamble SC et al. Discrete molecular classes of ovarian cancer suggestive of unique mechanisms of transformation and metastases. *Clin Cancer Res* 2014; 20: 87-99.
32. Ren H, Tan ZP, Zhu X et al. Identification of anaplastic lymphoma kinase as a potential therapeutic target in ovarian cancer. *Cancer Res* 2012; 72: 3312-3323.
33. Whiteside TL, Ferrone S. For breast cancer prognosis, immunoglobulin kappa chain surfaces to the top. *Clin Cancer Res* 2012; 18: 2417-2419.
34. Lundgren S, Berntsson J, Nodin B et al. Prognostic impact of tumour-associated B cells and plasma cells in epithelial ovarian cancer. *J Ovarian Res* 2016; 9: 21.
35. Schmidt M, Hellwig B, Hammad S et al. A comprehensive analysis of human gene expression profiles identifies stromal immunoglobulin kappa C as a compatible prognostic marker in human solid tumors. *Clin Cancer Res* 2012; 18: 2695-2703.
36. Warner JR, McIntosh KB. How common are extraribosomal functions of ribosomal proteins? *Mol Cell* 2009; 34: 3-11.

37. Loreni F, Mancino M, Biffo S. Translation factors and ribosomal proteins control tumor onset and progression: how? *Oncogene* 2014; 33: 2145-2156.
38. Doherty L, Sheen MR, Vlachos A et al. Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet* 2010; 86: 222-228.
39. Li C, Ge M, Chen D et al. RPL21 siRNA Blocks Proliferation in Pancreatic Cancer Cells by Inhibiting DNA Replication and Inducing G1 Arrest and Apoptosis. *Front Oncol* 2020; 10: 1730.
40. Suman S, Mishra A, Kulshrestha A. A systems approach for the elucidation of crucial genes and network constituents of cervical intraepithelial neoplasia 1 (CIN1). *Mol Biosyst* 2017; 13: 549-555.
41. Bedekovics T, Hussain S, Feldman AL, Galardy PJ. UCH-L1 is induced in germinal center B cells and identifies patients with aggressive germinal center diffuse large B-cell lymphoma. *Blood* 2016; 127: 1564-1574.
42. Ummanni R, Jost E, Braig M et al. Ubiquitin carboxyl-terminal hydrolase 1 (UCHL1) is a potential tumour suppressor in prostate cancer and is frequently silenced by promoter methylation. *Mol Cancer* 2011; 10: 129.
43. Ding X, Gu Y, Jin M et al. The deubiquitinating enzyme UCHL1 promotes resistance to pemetrexed in non-small cell lung cancer by upregulating thymidylate synthase. *Theranostics* 2020; 10: 6048-6060.
44. Liu S, Gonzalez-Prieto R, Zhang M et al. Deubiquitinase Activity Profiling Identifies UCHL1 as a Candidate Oncoprotein That Promotes TGFbeta-Induced Breast Cancer Metastasis. *Clin Cancer Res* 2020; 26: 1460-1473.
45. Gu Y, Lv F, Xue M et al. The deubiquitinating enzyme UCHL1 is a favorable prognostic marker in neuroblastoma as it promotes neuronal differentiation. *J Exp Clin Cancer Res* 2018; 37: 258.
46. Gutkin DW, Shurin MR, El Azher MA et al. Novel protein and immune response markers of human serous tubal intraepithelial carcinoma of the ovary. *Cancer Biomark* 2019; 26: 471-479.
47. Tangri A, Lighty K, Loganathan J et al. Deubiquitinase UCHL1 Maintains Protein Homeostasis through the PSMA7-APEH-Proteasome Axis in High-grade Serous Ovarian Carcinoma. *Mol Cancer Res* 2021; 19: 1168-1181.
48. Cucci MA, Grattarola M, Dianzani C et al. Ailanthone increases oxidative stress in CDDP-resistant ovarian and bladder cancer cells by inhibiting of Nrf2 and YAP expression through a post-translational mechanism. *Free Radic Biol Med* 2020; 150: 125-135.
49. Jin C, Yu W, Lou X et al. UCHL1 Is a Putative Tumor Suppressor in Ovarian Cancer Cells and Contributes to Cisplatin Resistance. *J Cancer* 2013; 4: 662-670.
50. Sahab ZJ, Hall MD, Me Sung Y et al. Tumor suppressor RARRES1 interacts with cytoplasmic carboxypeptidase AGL2 to regulate the alpha-tubulin tyrosination cycle. *Cancer Res* 2011; 71: 1219-1228.
51. Peterfi L, Banyai D, Yusenkov MV et al. Expression of RARRES1 and AGL2 and progression of conventional renal cell carcinoma. *Br J Cancer* 2020; 122: 1818-1824.

52. Huebner H, Strick R, Wachter DL et al. Hypermethylation and loss of retinoic acid receptor responder 1 expression in human choriocarcinoma. *J Exp Clin Cancer Res* 2017; 36: 165.

## Figures

### Figure 1

The technical road map.

### Figure 2

scRNA-seq data processing and analysis. A) The quality control chart of 9885 cells in GSE147082; B)

### Figure 3

Reconstruction of differentiation trajectories of ovarian cancer omentum metastasis site. A) The trajectory plot in 18 clusters using Monocle analysis; B) The trajectory plot of 14 cell types; C) The trajectory plot in Pseudotime; D) The trajectory plot of cells in 5 states.

### Figure 4

Identification of OMAGs in ovarian cancer omentum metastasis using scRNA-seq data and GEO data. A) K-M survival curve of 3 clusters, and the annual risk index of each cluster; B) Proportions of tumor stages in each cluster; C) Age distribution in each cluster; D) Distribution of anatomical sites in each cluster; E) Distribution of sample sources in each cluster; F) The expression of symbols of 5 branches in 3 clusters obtained from cell trajectory analysis, there was no statistically significant difference gene in branch 2.

### Figure 5

Tumor microenvironment and immune cell infiltration analysis. A) Violin maps showed the ESTIMATE Score, Immune Score, Stromal Score, Tumor Purity were significantly different between 3 clusters ( $***p < 0.001$ ); B) The proportion of 22 immune cells built on GEO microarray data; C) The difference of the fraction of 22 immune cells in 3 clusters ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ , ns: non-significance); D) The

Kaplan-Meier curves of patients with different proportions of T cells CD4 naïve, B cells naïve, macrophages M2, T cells CD4 memory activated, respectively.

## Figure 6

Analysis for several important immune checkpoints related with ovarian cancer. A) The 6 immune checkpoints with significant differences in expression among the 3 clusters were displayed by box plot (\*\*\*) ( $p < 0.01$ ); B-G) K-M survival curves were used to show the survival difference of high or low expression level of 6 immune checkpoint genes with significant difference in GEO data.

## Figure 7

Construct a OMAGs genetic risk model. A) The confidence interval under each lambda; B) The trajectory of each independent variable: the horizontal axis represents the log value of the independent variable lambda, and the vertical axis represents the coefficient of the independent variable. C) Effects of high and low risk scores of 6-OMAGs risk model on survival in the training set (TCGA-OV data); D) Effects of high and low risk scores of 6-OMAGs risk model on survival in the test set (GEO data); E) The prognostic value of the 6-OMAGs signature was evaluated using the ROC curves in training set (TCGA-OV data); F) The prognostic value of the 6-OMAGs signature was evaluated using the ROC curves in test set (GEO data).

## Figure 8

A) The univariate cox regression analyses of the 6-OMAGs and clinicopathological data; B) The multivariate Cox regression analysis of the 6-OMAGs and clinicopathological data; C) The bubble maps of the 6-OMAGs expression level in scRNA-seq samples of 18 clusters; D) The tSNE maps of expression of 6-OMAGs in scRNA-seq samples of 18 clusters. E) The nomogram model was constructed to predict the 1-, 3-, and 5-year survival of ovarian cancer patients.