Original Paper

Identification of Subpathway Signatures For Ovarian Cancer Prognosis by Integrated Analyses of High-Throughput miRNA and mRNA Expression

Songyu Tian^a Jiangtian Tian^{b,c} Xiuwei Chen^a Lianwei Li^a Yunduo Liu^a
Yuping Wang^d Yuqi Sun^e Chunlong Zhang^f Ge Lou^a

^aDepartment of Gynecology oncology, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang,

^bDepartment of Cardiology, The 2nd Affiliated Hospital of Harbin Medical university, Harbin,

Heilongjiang, ^cThe Key Laboratory of Myocardial Ischemia, Harbin Medical University, Ministry of

Education, Harbin, Heilongjiang, ^dDepartment of Gynecology, Yulin number 2 Hospital, Yulin city,

Shanxi, ^eDepartment of Gynecology and Obstetrics, Yichun Central Hospital, Yichun City, Heilongjiang,

^fCollege of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China

Key Words

Ovarian cancer • Prognosis • miRNA • Subpathway

Abstract

Background/Aims: Ovarian cancer (OC) causes more death and serious conditions than any other female reproductive cancers, and many expression signatures have been identified for OC prognoses. However, no significant overlap is found among signatures from different studies, indicating the necessity of signature identifications at the functional level. **Methods:** We performed an integrated analyses of miRNA and gene expressions to identify OC prognostic subpathways (pathway regions). Using The Cancer Genome Atlas data set, we identified core prognostic subpathways, and calculated subpathway risk scores using both miRNA and gene components. Finally, we performed global risk impact analyses to optimize core subpathways using the random walk algorithm. **Results:** Subpathway-level analyses displayed more robust results than the gene- and miRNA-level analyses. Moreover, we verified the advantage of core subpathways over the entire pathway-based results and their prognostic performance in two independent validation data sets. Based on the global impact score, 13 subpathway signatures were selected and a combined subpathway-based risk score was further calculated for OC patient prognoses. **Conclusions:** Overall, it was possible to systematically perform integrated analyses of the expression levels of miRNAs and genes to identify prognostic subpathways and infer subpathway risk scores for use in OC clinical applications.

© 2017 The Author(s)
Published by S. Karger AG, Basel

Chunlong Zhang
and Ge Lou

Department of Gynecology oncology, Harbin Medical University Cancer Hospital,
Harbin, College of Bioinformatics Science and Technology, Harbin Medical University,
Harbin, Heilongjiang (China); E-Mail zclbio1989@163.com, 158582518@qq.com

Introduction

Ovarian cancer (OC) is a widespread cancer that causes more deaths and serious conditions than any other female reproductive cancer. High-grade serous carcinoma is the major histological subtype (approximately 70%) of OC [1]. Female patients with OC may have no symptoms or mild symptoms until the cancer is in its advanced stages [2], which then responds poorly to treatment. Treatments for OC patients usually include surgery followed by chemotherapy and targeted therapy, according to the International Federation of Gynecology and Obstetrics (FIGO) staging system. However, the current FIGO staging system is not adequate. Even if OC patients have the same stage and similar clinical characteristics, they may experience different clinical outcomes. Due to the molecular heterogeneity of OC, some patients develop relapses, metastases, and drug resistances earlier than other patients. Gynecologists therefore need more specific OC tumor signatures that are superior to the FIGO staging system.

Bonome et al. used the Affymetrix human U133A microarray and Cox regression analyses to identify 57 signatures that were associated with OC survival [3]. Wang et al. identified 40 prognostic gene signatures in OC patients using the DirGenerank method [4]. However, even though these prognostic signatures possessed efficient power in their respective studies, there was no significant overlap among components in signatures. One possible reason for the poor overlap was that different components of signatures were involved in separate aspects of the same biological functions or pathways. Compared to gene-based signatures, function-based signatures could represent sets of genes with consistent functional roles that could result in a more robust performance. In addition, function-based analyses could resolve the dimensioned issues of transcriptomic data with more variables than the sample number [5, 6].

Recently, studies have analyzed tumor mechanisms at the functional or pathway levels. Ooi et al. developed an approach to identify the associations between molecular pathways and tumor profiles [7]. Similarly, Huang et al. developed a pathway-based model for survival prediction by combining Cox regression and functional information [8]. Another method, FAIME, was developed to generate “mechanism signatures” using rank-weighted gene expression levels derived from individual samples [6]. However, function-based analyses for identifying OC prognostic signatures has not yet been reported. More importantly, most of the above methods only considered gene expression to characterize function-level conditions, while ignoring the regulatory roles of non-coding RNAs.

MicroRNAs (miRNAs) are a class of non-coding RNAs that display key regulatory roles at the post-transcriptional level by binding to the 3'-untranslated regions of messenger RNAs (mRNAs) [9, 10]. The miRNAs were reported to be involved in tumor initiation, progression [11, 12] and prognosis [10]. For OC, some researchers also confirmed that miRNAs affected sensitivity of platinum [13] and tumor metastasis [14, 15]. Some studies have performed functional analyses (i.e., risk pathway identification) using miRNAs. In 2015, Kretschmann et al. performed pathway identification by counting miRNA numbers using the over-representation analysis strategy [16]. The integrated analyses of miRNA and mRNA expressions have also been used for many other purposes, including tumor biomarker identification [17], miRNA target optimization [18], and glioma mechanism analyses [19]. In our previous study, we constructed a glioma core survival network by simultaneously considering miRNA, mRNA expression and pathway topology, showing that the core survival network could effectively predict patient clinical outcomes [20]. It is therefore important to integrate miRNA and mRNA biomolecules for interpreting heterogeneous diseases, especially for tumor prognoses.

Biological pathways display advantages over gene ontology terms because of pathway topology structures. However, the large number of gene components within entire pathways present challenges for precise medical analyses. To resolve this issue, the subpathway (pathway regions) was defined based on pathway topology [21]. Compared with entire pathway, the subpathway contained a small set of components and reflected detailed

functional conditions. Furthermore, it has been reported that abnormalities of subpathways, and not entire pathway, were involved in disease etiology [21, 22]. Recently, subpathway-level analyses were also applied to systematically characterize mechanisms of drug actions [23] and miRNA regulation [24]. In 2015, we performed integrated analyses to identify a cell cycle subpathway as a signature for the prognoses of lung cancer patients [25]. To further expand the roles of miRNAs when performing subpathway-level analyses, our laboratory included miRNA molecules into the pathway topology structure to develop an available R package (*subPathwayGMir*) for miRNA and mRNA integrated analyses [26].

In the present study, we comprehensively integrated high-throughput miRNA and gene expressions to identify OC survival-related subpathways and to calculate subpathway risk scores for patient prognoses. Using The Cancer Genome Atlas (TCGA) data set, we found that the subpathway-level results were more robust than the gene- and miRNA-level results. Regarding prognostic performance, the core subpathways displayed more predictive power than core pathways. Finally, we identified a 13 subpathway signatures using risk impact analyses based on global pathway networks, and confirmed independent predictive power of the subpathway signature using two validation data sets.

Materials and Methods

Datasets

Training data set from the TCGA. The OC data set, including molecular data and patient clinical information, was downloaded from the TCGA portal. For the molecular data, the gene expression was generated using HTseq-FPKM, and miRNA expression was generated using BCGSC miRNA Profiling. The average expression values were calculated for duplicated samples. Moreover, we eliminated samples with survival times < 30 days, because these patients might have died for reasons other than the disease [27]. In this study, the expression data and clinical information of a total of 370 patients were used as the training set. For comparative analyses, we further randomly divided this data set into two subsets (TCGA set 1 and TCGA set 2), which contained the same number of patients. The patient clinical features of these two subsets are listed (for all online suppl. material, see www.karger.com/doi/10.1159/000485492) in Suppl. Table 1.

Validation data sets. We obtained two available validation data sets to test the prognostic performance of subpathway signatures. The validation data set 1 was downloaded from the study of Korsunsky et al. from Gene Expression Omnibus (GEO) database [28]. The access number of gene expression was GSE81778 and the access number of miRNA expression was GSE81873. Finally, the expression data sets and clinical information of 19 patients were utilized. We also obtained validation data set 2 from International Cancer Genome Consortium (ICGC) database (<http://icgc.org/>). Within this data set, a total of 93 OC tumor samples with available miRNA, gene expression, and clinical information were analyzed.

Biological pathway information

The information regarding biological pathways was derived from the Kyoto Encyclopedia of Genes and Genomes PATHWAY database. We applied Bioconductor package *SubpathwayGMir* [26] to obtain all biological pathways, including 152 metabolic pathways and 191 non-metabolic pathways, and utilized these pathways to identify OC survival-related subpathways.

Subpathway identifications and risk score calculations

We identified survival-related subpathways by integrated analyses of survival-related miRNAs and genes. The detailed procedure was as follows:

- i) *Identification of survival related miRNAs and genes.* First, we identified the survival-related miRNAs and genes using the univariate Cox regression.
- ii) *Locate subpathways mediated by miRNAs.* We mapped the survival-related miRNAs and genes as signature nodes into the reconstructed pathway graphs obtained from the *SubpathwayGMir* package [26]. We then computed the shortest path between any two signatures, and two parameters (n and s), controlling the intensity of signature signals and the size of candidate subpathways. The default parameters, $n = 5$ and $s = 5$, were used in the calculations.

iii) *Evaluation of the statistical significance.* To evaluate whether the located subpathways were associated the patient prognoses, we further utilized the hypergeometric method to calculate the statistical significance as follows:

In the formula, m_g (m_{mir}) was the number of genes (miRNAs) in the entire genome (miRNAome), and n_g

$$P = 1 - \sum_{k=0}^{r_g + r_{mir} - 1} \frac{\binom{t_g + t_{mir}}{k} \binom{m_g + m_{mir} - t_g - t_{mir}}{n_g + n_{mir} - k}}{\binom{m_g + m_{mir}}{n_g + n_{mir}}}$$

(n_{mir}) was the number of survival related genes (miRNAs), of which r_g (r_{mir}) genes (miRNAs) were included in the subpathway region containing t_g genes (t_{mir} miRNAs). The adjusted P-values were calculated using the Benjamini-Hochberg method.

For each subpathway, we calculated risk score by considering miRNA and gene components. The score was a linear combination of expression values of each miRNA and gene in the subpathway multiplied by weighted values estimated using univariate Cox proportional hazards regression. Based on the risk score, a subpathway-based matrix with subpathways as rows and OC samples as columns was developed. We respectively identified prognostic subpathways using two TCGA subsets and the common subpathways between two subsets that were derived from the same entire pathway were regarded as core subpathways. Furthermore, we merged these core subpathways if they satisfied the following formula:

$$|P_i \cap P_j| > 0.8 * \min\{|P_i|, |P_j|\}$$

In the formula, the P_i and P_j were two subpathways from the same entire pathway.

Global risk impact evaluations

Based on core subpathways, we performed global risk impact analyses to optimize these subpathways using random walk algorithm [29]. First, we merged all biological pathways to construct a global pathway network that contained miRNA and gene components. Then, we annotated survival-related miRNAs and genes within these core subpathways into the global network, and regarded these miRNAs and genes as seed nodes. Finally, random walk algorithm was used to evaluate the global risk impact of seed nodes on all miRNA and gene components as follows:

$$P^{t+1} = (1-r)WP^t + rP^0$$

where W was the column-normalized adjacency matrix of the global pathway network, which consisted of 0 and 1. P^t was a vector in which a node in the global network held the probability of finding itself in process up to step t . In this study, the initial probability vector, P^0 , was constructed in such a way that equal probabilities were assigned to all seed nodes, and the sum of their probabilities was equal to 1. Additionally, the restart of walker at each step was r ($r = 0.7$). When the difference between P^t and P^{t+1} fell below 10^{-6} , the probabilities reached a steady state. Then, all miRNA and gene components obtained their scores according to the values in steady-state probability vector, P^∞ .

Survival analyses

We utilized Cox proportional hazards regression to identify survival-related miRNAs and genes based on corresponding high-throughput data sets. For subpathway risk matrix, K-mean clustering method ($K = 2$) was used to divide tumor samples into two risk groups, low-risk and high-risk, according to the subpathway risk score. Based on the 13 subpathway signatures, a Combined subPathway-Based Risk (CsPBR) score was calculated as the average value of these 13 subpathway risk scores. For each OC sample, a CsPBR score was assigned and the median CsPBR score was used as cut-off for OC sample classifications. The Kaplan-Meier method was used to generate survival curves and the difference between survival curves was evaluated using the log-rank test. All tests were two-tailed, and a P-value < 0.05 was considered a significant result.

Results

Subpathway-level results were more robust than miRNA- and gene-level results.

To perform robust analyses of different level results, we first randomly divided the 370 samples from TCGA training data set into two subsets, TCGA set 1 (n = 185) and TCGA set 2 (n = 185). The results showed that there was no significant difference between the clinical characteristics of these two sample sets (see online suppl. material, Suppl. Table 1). In the TCGA set 1, we identified a total of 1,511 survival-related genes and 95 survival-related miRNAs using univariate Cox analyses with a value of P-value < 0.05, and 82 survival-related subpathways were identified using integrated analyses of survival-related miRNAs and genes with adjusted P-value < 0.05 (see Materials and Methods; Fig. 1A). In the TCGA set 2, we identified 2,104 survival-related genes and 132 survival-related miRNAs with P-value < 0.05, and 140 prognostic subpathways with adjusted P-value < 0.05.

To compare the robustness of the subpathway-level and miRNA- or gene-level results, we evaluated the component consistencies for different levels using these two TCGA subsets. The significance of common components in each level assumed that even with biological variabilities and patient heterogeneities, the higher consistencies between different subsets indicated a higher quality in the signature selection. Fig. 1B shows that at the subpathway level, the 82 subpathways from set 1 and 140 subpathways from set 2 entirely shared a total of 61 pathways, and the P-value calculated by the hypergeometric test was 1.1×10^{-16} (pathway background number: 343). For example, subpathway path: 00010_1 from set 1 and subpathway path: 00010_2 from set 2 were derived from the same entire pathway (path: 00010). However, at the miRNA and gene levels, poor consistencies between components were observed; only 10 miRNAs were shared by 95 miRNAs from set 1 and 132 miRNAs from

Fig. 1. A comparison of the robustness of the subpathway-, miRNA- and gene-levels. (A) The analytical procedure for performing the robustness comparison. (B) Comparison between the miRNA level, gene level and subpathway level using different cutoff values. The final P-values were calculated using the hypergeometric test, and the background numbers for the miRNA, gene and pathway were 1881, 30354 and 343, respectively.

set 2 (P-value = 0.12; miRNA background number: 1881), and 104 genes were shared by 1, 511 genes from set 1 and 2, 104 genes from set 2 (P-value = 0.55; gene background number: 30354). When setting more strict cut-off values for identifying subpathways, we also observed significant robust results at the subpathway level (P-value = $8.9\text{E-}16$ for adjusted P-value < 0.01 and P-value = $1.1\text{E-}12$ for adjusted P-value < 0.001; Fig. 1B). Similarly, we set different cut-off values to identify the survival-related miRNAs and genes. Only 14 gene components were significantly shared between the two TCGA subsets when P-value < 0.01 at the gene level. By analyzing the consistencies of signatures, these results confirmed the subpathway-level results were more robust than the gene- or miRNA-level results. In this study, the subpathways (pathways), which were shared by two TCGA subsets, were regarded as the core subpathways (pathways).

Core subpathways predict patient clinical outcomes in the TCGA data set.

For each of the core subpathways, a risk score was calculated for each tumor sample by considering all the miRNA and gene components in this subpathway, which included survival-related miRNAs and genes and non-survival miRNAs and genes (see Materials and Methods). We identified core subpathways using one of the TCGA subsets, and tested the predictive performance using another subset. This process involved a subpathway risk matrix formed by calculating the risk scores using Cox weighting values from one TCGA subset (inner training set) and the expression values from another subset (inner testing set). Based on the risk score matrix, we performed K-mean clustering to classify tumor samples from the inner testing set into high-risk and low-risk groups, and the log-rank method was used to calculate significant difference of survival times between these two risk groups. Suppl. Fig. 1A and 2B (see online suppl. material) show that the core subpathways that were identified by an adjusted P-value < 0.05 significantly predicted patient clinical outcomes (log-rank P-value = 0.013 in TCGA set 1 and log-rank P-value = 0.031 in TCGA set 2). When considering more

Fig. 2. A comparison of prognostic performances. (A) Comparisons of the signature prognostic ability in the Cancer Genome Atlas subsets. (B) Comparisons between core subpathway signatures and random results. Left panel, the signatures obtained using an adjusted P-value < 0.05; middle panel, the signatures obtained using an adjusted P-value < 0.01; right panel, the signatures obtained using an adjusted P-value < 0.001. The red and green circles designate the results of the subpathway signatures and pathway signatures, respectively. The grey circles designate the random results.

Fig. 3. The distribution for global impact scores of (A) miRNA and gene components and (B) 106 core subpathways.

stringent cutoff values (0.01 and 0.001) to identify core subpathways, we also observed that the core subpathways were significantly associated with OC patient survival (see online suppl. material, Suppl. Fig. 1E, 1F, 1I and 1J). To further test the robustness of the survival results, we perturbed the patients' survival times and developed 1, 000 random clinical data sets. Again, we performed survival analyses for the core subpathways based on these random clinical data sets. Finally, their prognostic performances were evaluated with true results in both TCGA subsets by using the negative log-transformed P-values of the log-rank test. The higher the two scores, the better the prognostic signature performance. Fig. 2A and 2B show that the core subpathways displayed better predictive performance in the true data set than these random data sets; only 6, 3, and 2 cases of 1, 000 random analyses displayed better predictive results. In conclusion, these findings showed that the core subpathways were associated with OC patient prognoses.

Prognostic performance comparisons between core subpathways and pathways

Similar to the subpathway level, we also performed survival analyses to test whether corresponding entire pathways could significantly predict OC patient clinical outcomes. First, using the two TCGA subsets, we calculated the risk scores for each of core pathways by considering all the miRNA and gene components involved in the entire pathway. Notably, the entire pathway contained more components than the corresponding subpathways. Then, based on the pathway risk matrix, the K-mean clustering method was used to divide tumor samples into low-risk and high-risk groups. Finally, the log-rank test was used to calculate the statistical significance of survival difference between these two risk groups. Suppl. Fig. 1C and 1D (see online suppl. material) show that the 61 entire pathway (with adjusted P-value < 0.05) did not show prognostic significance in the two TCGA subsets (P-value = 0.449 in set 1 and P-value = 0.201 in set 2). Consistent survival results were also observed in the 50 and 31 core pathways with less adjusted P-values (see online suppl. material, Suppl. Fig. 1G, 1H, 1K and 1L). Moreover, the prognostic performance of the entire pathways was similar with the random results (Fig. 2A), reflecting the advantage of subpathway-level analyses over the entire pathway analyses.

Core subpathways predict patient clinical outcomes in two validation data sets.

For the core subpathways with adjusted P-value < 0.05, we further merged the subpathways that were derived from the same entire pathway and shared enough components (see Materials and Methods). The detailed workflow is shown in see online suppl. material, Suppl. Fig. 2. As a result, a total of 106 core subpathways were identified

based on 62 subpathways from the TCGA set 1 and 65 subpathways from the TCGA set 2. To further test the prognostic values of these subpathways, we utilized the total TCGA data set as training set, and the two data sets from GEO and ICGC as testing sets. Similarly, the risk score was calculated for these 106 core subpathways, and a subpathway matrix with subpathways as rows and OC samples as columns was formed for validation data sets. In the GEO validation data set, the 106 core subpathways were not significantly associated with patient clinical outcomes (P-value = 0.656; see online suppl. material, Suppl. Fig. 3A). These core subpathways were significantly associated with patient clinical outcomes with P-value = 1.46×10^{-4} in the ICGC validation data set (see online suppl. material, Suppl. Fig. 3B). We also tested whether the miRNA and gene components within these subpathways were associated with patient survival. We performed the K-mean clustering method for validation data sets by utilizing the expression levels of miRNAs and genes, and the log-rank was used to calculate the survival difference. Suppl. Fig. 3C and 3D (see online suppl. material) show that the miRNA and gene components within core subpathways also significantly predicted OC patient clinical outcomes with P-value = 0.048 and 8.3×10^{-5} in the two validation data sets.

Thirteen subpathway signatures were identified by global risk impact analyses.

To further optimize the 106 core subpathways, we performed a global risk strategy using random walk method based on global pathway topology structure (see Materials and Methods). During random walk process, the survival-related miRNAs and genes within core subpathways were regarded as seed nodes, and all miRNA and gene components were assigned a global impact score based on the final steady-state probabilities. The impact score reflected the associations of each component and seed nodes which represented prognostic signals in the global pathway network. Because the scores of most components were very similar, the scores were further $\log_{10}$ transformed and regarded as the global impact (GI) score (the more negative of GI score, the less significant). The GI score for core subpathways was defined by calculating the average GI score of all miRNA and gene components within this subpathway. Fig. 3 shows the GI score distribution for the miRNA (gene) components and all core subpathways. The top 13 subpathways with high GI scores were defined as prognostic signatures (see Fig. 4).

Based on these 13 subpathways, we further defined a CsPBR score for each OC patient, which indicated the average risk score of these key subpathways (see Materials and Methods). We further tested the prognostic performance of the 13- subpathway signatures

Fig. 4. The 13 subpathway signature results. (A) The subpathway information. (B) The subpathway networks. The triangles designate miRNAs and the rectangles designate genes. The node color designates the global impact scores for the miRNA and gene components.

Fig. 5. The Kaplan-Meier curves of the 13 subpathway signatures using the combined subpathway-based risk scores in the (A) Gene Expression Omnibus validation data set and (B) the International Cancer Genome Consortium validation data set.

using two validation data sets. Assigning the median CsPBR value as the cutoff, we divided the OC samples from validation data set into two risk groups. The log-rank method was used to calculate the survival difference between different risk groups. In the GEO data set, the median survival time of patients with higher CsPBR scores was 29.5 months, whereas the median survival time of patients with lower CsPBR scores was 64 months. The log-rank result showed that the subpathway signature was significantly associated with patient prognoses with log-rank P-value = 0.013 (see Fig. 5A). In the ICGC data set, the median survival time of patients with higher CsPBR scores was 822.5 days, whereas the median survival time of patients with lower CsPBR scores was 1,527 days. Similarly, the log-rank result showed that the subpathway signature was also significantly associated with patient prognoses with log-rank P-value = 0.025 (see Fig. 5B).

Discussion

For certain types of tumors, the gene components of signatures derived from different studies often poorly overlap, which emphasizes the need for a function-based strategy [5, 6]. In the present study, we first identified the survival-related miRNAs, genes and subpathways using two TCGA subsets. By comparing the consistency of signatures, we found that the function-level results were more robust than the miRNA- and gene-level results, which was consistent with previous findings. In addition, we used both miRNA and gene expressions in the subpathway identification based on the reconstructed pathway platform containing

miRNA molecules. This is the first study done to identify OC prognostic subpathways by integrating both miRNA and gene components.

Based on these survival-related subpathways, we further calculated risk scores by combining Cox coefficient and expression values of all miRNA and gene components involved in corresponding subpathway. Notably, the survival-related subpathways were identified from survival-related miRNAs and genes using hypergeometric test; however, all components within the subpathways were included in the risk score calculations. The risk score therefore reflected the average risk conditions at the subpathway level. During the risk score calculations, we utilized different data sets to obtain the Cox efficient and expression values. As an example, the subpathway risk score for TCGA set 2 was calculated by combining the Cox efficient from TCGA set 1 (inner training set) and the expression values from TCGA set 2 (inner testing set) based on all miRNAs and genes. In the validation verification, the TCGA data set was treated as training set and the validation data sets as testing sets.

In the prognostic performance comparison, we observed that the core subpathways displayed better predictive power than the entire pathways, which reflected the advantage of subpathway-level analyses. In the survival analyses, clustering method was used to divide all analytical samples into high-risk and low-risk groups based on the subpathway risk scores. However, the sample number was not balanced between these two risk groups obtained by K-mean clustering, and the 106 core subpathways were too many for a specific medical diagnoses. We therefore performed global impact analyses to select the higher risk subpathways as prognostic signatures, and developed CsPBR scores for dividing tumor samples, which overcame two issues mentioned above. We considered two aspects, enrichment analyses by survival-related miRNAs (and genes) and global impact scores for all miRNA (and gene) components, to identify the 13 subpathway signatures.

Among these 13 subpathways, one subpathway (path: 04666_2) from Fc gamma R-mediated phagocytosis ranked first with a GI score of -2.702. It has been confirmed that genes from Fc gamma R-mediated phagocytosis were abnormally expressed in breast cancer, which was another type of female tumor [30]. And the mTOR signaling pathway was often dysregulated in OC patients [31]. Moreover, some cancer subpathways, including renal cell carcinoma (path 05211_1), prostate cancer (path 05215_2), and pancreatic cancer (path 05212_1) were also identified, suggesting the presence of similar pathological mechanisms among these tumors and OC. In addition, some miRNA and gene components within these subpathways were also involved in the OC biological mechanism. For example, Ling et al. speculated that miR-451 was a novel prognostic marker, and the regulatory roles of miR-143 in the epithelium of OC patients have been confirmed [32, 33]. And these two miRNAs were both included in the subpathway (path: 04666_2) with highest GI score.

We integrated high-throughput miRNA, mRNA expression, and pathway structure to systematically identify a 13 subpathway signatures, and define CsPBR scores for predicting OC patient clinical outcomes. Sample-matched miRNA and mRNA expression data set with patient survival data has recently been developed, and further survival verification for 13 subpathway signatures using more data sets was necessary. Overall, the results of the present study are potentially useful for understanding the mechanism of OC, and identifying functional signatures for patient clinical prognoses.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (Grant. 81472028), Harbin Medical University Cancer Hospital Haiyan scientific research fund (JJZD2017-04), Harbin city technology bureau (2015RAXYJ054) and Harbin medical university scientific research innovation fund (2016JCZX48).

Disclosure Statement

The authors declare that there are no conflicts of interest.

References

- 1 Darb-Esfahani S, Kunze CA, Kulbe H, Sehoul J, Wienert S, Lindner J, Budczies J, Bockmayr M, Dietel M, Denkert C, Braicu I, Johrens K: Prognostic impact of programmed cell death-1 (pd-1) and pd-ligand 1 (pd-l1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high grade serous carcinoma. *Oncotarget* 2016;7:1486-1499.
- 2 Ebell MH, Culp MB, Radke TJ: A systematic review of symptoms for the diagnosis of ovarian cancer. *Am J Prev Med* 2016;50:384-394.
- 3 Bonome T, Levine DA, Shih J, Randonovich M, Pise-Masison CA, Bogomolny F, Ozbun L, Brady J, Barrett JC, Boyd J, Birrer MJ: A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res* 2008;68:5478-5486.
- 4 Wang JY, Chen LL, Zhou XH: Identifying prognostic signature in ovarian cancer using dirgengerank. *Oncotarget* 2017;8:46398-46413.
- 5 Ma S, Kosorok MR, Huang J, Dai Y: Incorporating higher-order representative features improves prediction in network-based cancer prognosis analysis. *BMC Med Genomics* 2011;4:5.
- 6 Yang X, Regan K, Huang Y, Zhang Q, Li J, Seiwert TY, Cohen EE, Xing HR, Lussier YA: Single sample expression-anchored mechanisms predict survival in head and neck cancer. *PLoS Comput Biol* 2012;8:e1002350.
- 7 Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, Ward L, Koo JH, Gopalakrishnan V, Zhu Y, Cheng LL, Lee J, Rha SY, Chung HC, Ganesan K, So J, Soo KC, Lim D, Chan WH, Wong WK, Bowtell D, Yeoh KG, Grabsch H, Boussioutas A, Tan P: Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *PLoS Genet* 2009;5:e1000676.
- 8 Huang S, Yee C, Ching T, Yu H, Garmire LX: A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer. *PLoS Comput Biol* 2014;10:e1003851.
- 9 Rottiers V, Naar AM: Micrnas in metabolism and metabolic disorders. *Nat Rev Mol Cell Biol* 2012;13:239-250.
- 10 Tomasetti M, Santarelli L, Neuzil J, Dong L: Microna regulation of cancer metabolism: Role in tumour suppression. *Mitochondrion* 2014;19 Pt A:29-38.
- 11 Teng Y, Zuo X, Hou M, Zhang Y, Li C, Luo W, Li X: A double-negative feedback interaction between microrna-29b and dnmt3a/3b contributes to ovarian cancer progression. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2016;39:2341-2352.
- 12 Ying X, Wei K, Lin Z, Cui Y, Ding J, Chen Y, Xu B: Microrna-125b suppresses ovarian cancer progression via suppression of the epithelial-mesenchymal transition pathway by targeting the set protein. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2016;39:501-510.
- 13 Chen Q, Qin R, Fang Y, Li H: Berberine sensitizes human ovarian cancer cells to cisplatin through mir-93/pten/akt signaling pathway. *Cell Physiol Biochem* 2015;36:956-965.
- 14 Li P, Sun Y, Liu Q: Microrna-340 induces apoptosis and inhibits metastasis of ovarian cancer cells by inactivation of nf-x03ba;b1. *Cell Physiol Biochem* 2016;38:1915-1927.
- 15 Song N, Liu H, Ma X, Zhang S: Placental growth factor promotes metastases of ovarian cancer through mir-543-regulated mmp7. *Cell Physiol Biochem* 2015;37:1104-1112.
- 16 Kretschmann A, Danis B, Andonovic L, Abnaof K, van Rikxoort M, Siegel F, Mazzuferi M, Godard P, Hanon E, Frohlich H, Kaminski RM, Foerch P, Pfeifer A: Different microrna profiles in chronic epilepsy versus acute seizure mouse models. *J Mol Neurosci* 2015;55:466-479.
- 17 Van der Auwera I, Limame R, van Dam P, Vermeulen PB, Dirix LY, Van Laere SJ: Integrated mirna and mrna expression profiling of the inflammatory breast cancer subtype. *Br J Cancer* 2010;103:532-541.

- 18 Zhu M, Yi M, Kim CH, Deng C, Li Y, Medina D, Stephens RM, Green JE: Integrated mirna and mrna expression profiling of mouse mammary tumor models identifies mirna signatures associated with mammary tumor lineage. *Genome Biol* 2011;12:R77.
- 19 Dong H, Siu H, Luo L, Fang X, Jin L, Xiong M: Investigation gene and microRNA expression in glioblastoma. *BMC Genomics* 2010;11:S16.
- 20 Zhang C, Li C, Li J, Han J, Shang D, Zhang Y, Zhang W, Yao Q, Han L, Xu Y, Yan W, Bao Z, You G, Jiang T, Kang C, Li X: Identification of mirna-mediated core gene module for glioma patient prediction by integrating high-throughput mirna, mrna expression and pathway structure. *PLoS One* 2014;9:e96908.
- 21 Li C, Li X, Miao Y, Wang Q, Jiang W, Xu C, Li J, Han J, Zhang F, Gong B, Xu L: Subpathwayminer: A software package for flexible identification of pathways. *Nucleic Acids Res* 2009;37:e131.
- 22 Li C, Han J, Yao Q, Zou C, Xu Y, Zhang C, Shang D, Zhou L, Sun Z, Li J, Zhang Y, Yang H, Gao X, Li X: Subpathway-gm: Identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. *Nucleic Acids Res* 2013;41:e101.
- 23 Li C, Shang D, Wang Y, Li J, Han J, Wang S, Yao Q, Zhang Y, Zhang C, Xu Y, Jiang W, Li X: Characterizing the network of drugs and their affected metabolic subpathways. *PLoS One* 2012;7:e47326.
- 24 Li X, Jiang W, Li W, Lian B, Wang S, Liao M, Chen X, Wang Y, Lv Y, Yang L: Dissection of human mirna regulatory influence to subpathway. *Brief Bioinform* 2012;13:175-186.
- 25 Zhang C, Li C, Xu Y, Feng L, Shang D, Yang X, Han J, Sun Z, Li Y, Li X: Integrative analysis of lung development-cancer expression associations reveals the roles of signatures with inverse expression patterns. *Mol Biosyst* 2015;11:1271-1284.
- 26 Feng L, Xu Y, Zhang Y, Sun Z, Han J, Zhang C, Yang H, Shang D, Su F, Shi X, Li S, Li C, Li X: Subpathway-gmir: Identifying mirna-mediated metabolic subpathways by integrating condition-specific genes, microRNAs, and pathway topologies. *Oncotarget* 2015;6:39151-39164.
- 27 Srinivasan S, Patric IR, Somasundaram K: A ten-microRNA expression signature predicts survival in glioblastoma. *PLoS One* 2011;6:e17438.
- 28 Korsunsky I, Parameswaran J, Shapira I, Lovecchio J, Menzin A, Whyte J, Dos Santos L, Liang S, Bhuiya T, Keogh M, Khalili H, Pond C, Liew A, Shih A, Gregersen PK, Lee AT: Two microRNA signatures for malignancy and immune infiltration predict overall survival in advanced epithelial ovarian cancer. *J Investig Med* 2017;65:1068-1076.
- 29 Kohler S, Bauer S, Horn D, Robinson PN: Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet* 2008;82:949-958.
- 30 Peng Q, Zhao L, Hou Y, Sun Y, Wang L, Luo H, Peng H, Liu M: Biological characteristics and genetic heterogeneity between carcinoma-associated fibroblasts and their paired normal fibroblasts in human breast cancer. *PLoS One* 2013;8:e60321.
- 31 Levine DA, Bogomolny F, Yee CJ, Lash A, Barakat RR, Borgen PI, Boyd J: Frequent mutation of the pik3ca gene in ovarian and breast cancers. *Clin Cancer Res* 2005;11:2875-2878.
- 32 Ling S, Ruiqin M, Guohong Z, Ying W: Expression and prognostic significance of microRNA-451 in human epithelial ovarian cancer. *Eur J Gynaecol Oncol* 2015;36:463-468.
- 33 Wang L, He J, Xu H, Xu L, Li N: Mir-143 targets ctgf and exerts tumor-suppressing functions in epithelial ovarian cancer. *Am J Transl Res* 2016;8:2716-2726.