mRNAsi Index-based Integrative Analysis for the Identification of New Prognostic Models for Ovarian Cancer

Xue Pan
Shengjing Hospital of China Medical University

Xiaoxin Ma (maxiaoxin12345678@163.com)
Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China. https://orcid.org/0000-0002-4147-3651

Research article

Keywords: ovarian cancer, mRNAsi, ceRNA, immune infiltration, machine learning, prognosis

DOI: https://doi.org/10.21203/rs.3.rs-76966/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Ovarian cancer (OC) has the highest mortality rate among all female reproductive system malignant tumors worldwide. In this study, we aimed to investigate OC from several perspectives by using machine learning. Our results showed that the mRNA expression-stemness index (mRNAsi) is closely related to clinical characteristics of OC patients, as OC patients with venous or lymphatic invasion had higher mRNAsi score compared to patients with no invasion. Further grade 3/4 patient group had higher mRNAsi scores compared to the grade1/2 group. We also found that mRNAsi is closely related to immune infiltration in OC. We also built a competing endogenous RNA network, which contained 4 miRNAs, 5 lncRNAs, and 1 mRNA, by using Cytoscape based on the differentially expressed genes of the high- and low-mRNAsi groups. Through Lassio regression, we also established a model including 7 lncRNAs and 2miRNAs, which could effectively categorize OC patients into two groups based on the median risk score. We then developed a nomogram model which could effectively forecast the overall survival rate of OC for 1-, 3-, and 5-year period. The models assessed in this study showed potential for clinical application in treatment decisions for OC.

Introduction

Ovarian cancer (OC), a “silent killer”, is the deadliest female reproductive system cancer [1, 2] as well as the second most common gynecological malignant tumor in females over 40 worldwide. OC incidence in developing countries in Asia and the Middle East is a little lower than it in the developed countries in the North America and Europe [3]. At diagnosis, OC is usually almost at an advanced stage with poor prognosis. Over 70% of OC patients are at stage III or IV at the time of diagnosis [2, 4], and this is due to the absence of particular symptoms as well as early metastasis [5, 6]. The overall long-term prognosis of advanced OC has remained poor over the last 30 years [2, 7] despite the ongoing development of both surgical and chemotherapy treatment. The 5-year overall survival rate of patients diagnosed at stage I is 90%, at stage II-65%, at stage III-34%, and stage IV-15% [8]. In addition, OC has high recurrence, with 75% of patients with advanced OC experiencing a relapse within 3 years and follow-on chemoresistance making OC treatment even more difficult. Various research has described the significant role of cancer stem cells (CSCs) in the recurrence, metastasis, heterogeneity, and multidrug resistance of OC [9].

CSCs were first identified and isolated from leukemia in 1994 [10, 11]. Since then, they have been the subject of various investigations worldwide. CSCs are described as the “heart” of cancers, as they have the capacity of self-renewal and can differentiate into cancer cell lineages [12, 13]. In 2005, Bapat and colleagues found that CSCs also exist in OC [14]. Subsequent studies revealed that CSCs have significant roles in the initiation, metastasis, and chemoresistance of OC. However, despite extensive studies, various aspects of these roles of CSCs remain unclear. Thus, a better understanding of the nature of OC stem cells could contribute to the development of new treatment strategies. Tumor-infiltrating immune cell (TIICs), as members of the tumor microenvironment, also play a critical role in the pathogenesis of OC [15, 16]. In general, there are 22 functional immune cell types including monocytes, B cells naïve, M1
macrophages, CD8 T cells, natural killer cells (NK cells), and others. Whether there is a relationship between OC stemness and tumor immune cell infiltration remains ambiguous.

Machine learning can be applied not only for search engines and speech recognition [17], but also in molecular biology. Unsupervised learning, a kind of machine learning, is widely used as a bioinformatics method for medical data mining. The one-class logistic regression (OCLR), a type of unsupervised machine learning introduced in 2016 [18], is used to study the features of cell subtypes by deconvolution of high-throughput tumor biopsy data. The mRNA expression-stemness index (mRNAsi) quantifies the characterization of CSCs through the extraction of transcriptomic features from pluripotent stem cell samples via the OCLR algorithm. In general, higher mRNAsi scores indicate malignant biological processes and a poor prognosis. Novel biomarkers of OC stem cells can be identified for mRNAsi through the use of bioinformatics. In order to relate this to tumor immune infiltration, CIBERSORT algorithm can be employed to infer the relative proportions of different kinds of tumor-infiltrating immune cells.

Long non-coding RNAs (lncRNA) exert essential effects for the initiation and progression of tumors. lncRNA competes with endogenous RNAs through binding to microRNA (miRNA) and then regulates the expression of certain mRNA transcripts. Although there have been several studies on the role of lncRNAs in OC [19–21], a competing endogenous RNA (ceRNA) network based on OC stemness has not yet been reported.

In the current study, OC cases downloaded from The Cancer Genome Atlas (TCGA) were divided into two groups, namely OC with a high and a low mRNAsi score (high-mRNAsi and low-mRNAsi, respectively). A ceRNA network was built on the basis of differentially expressed miRNAs, IncRNAs, and mRNAs selected by comparing the two groups (high-mRNAsi vs low-mRNAsi). The correlation of copy number variation of differentially expressed mRNAs and the mRNAsi score was then analyzed. The prognostic value of differentially expressed miRNAs, IncRNAs, and mRNAs was assessed by utilizing univariate cox proportional hazards regression analysis (Cox PHR). The IncRNAs, miRNAs, and mRNAs associated with prognosis were fitted into LASSO regression to establish the risk regression model. In addition, a nomogram model based on the selected prognosis-related genes was built to facilitate clinical application. The clinical value of the nomogram model was evaluated through a calibration curve, DCA curve, and clinical impact curve. The immune infiltration profile of the OC data was also analyzed to further investigate the relationship between immune infiltration profile and the mRNAsi of OC.

Materials And Methods

Data collection

Transcriptome profiling data, copy number variation data, and clinical characteristic of OC cases were obtained from the TCGA database (https://portal.gdc.cancer.gov/). Transcriptome profiling data included the expression of lncRNAs, miRNAs, and mRNAs. Clinical characteristic included survival time, life-state,
age, stage, grade, venous invasion, and lymphatic invasion. Transcriptome profiling lncRNAs and mRNAs data included 379 OC samples, and transcriptome profiling miRNAs data included 499 samples.

**Analysis of the relationship between mRNAsi and clinical characteristics**

mRNAsi quantified the characterization of CSCs through extraction of the transcriptomic features from pluripotential stem cell samples (embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs)) based on the same method of Malta et al through use of the OCLR algorithm [22]. The Wilcoxon rank-sum test was used to analyze the relationship between mRNAsi and clinical characteristics including age, stage, grade, venous invasion and lymphatic invasion. A Kaplan–Meier (K-M) curve was plotted to estimate the prognostic value of mRNAsi for OC.

**Identification of differentially expressed, miRNAs, lncRNAs, and mRNAs related to mRNAsi**

Transcriptome profiling data was divided into two groups (high- and low-mRNAsi group) on the basis of median mRNAsi score. The “limma” package was used to screen for differentially expressed miRNAs, lncRNAs, and mRNAs. The miRNAs, lncRNAs, and mRNAs with false discovery rate (FDR) < 0.05 and |log2 (fold change)| > 1 were considered statistically significant. Finally, a volcano plot was drawn to visualize results.

**Construction of the lncRNA-miRNA-mRNA ceRNA network**

The selected differentially expressed miRNAs, lncRNAs, and mRNAs were used for the construction of a network based on Pearson correlation analysis. The network was visualized by using the Cytoscape software. Subsequently, we refined and simplified the network diagram according to the principle of ceRNA (lncRNA and mRNA have a negative correlation with miRNA, IncRNA has a positive correlation with mRNA). This simplified network was considered as the final ceRNA network.

**Analysis of the correlation between the mRNAsi and copy number variation of differentially expressed mRNAs**

Classically, copy number variation (CNV) refers DNA fragments that have been amplified or reduced, resulting in genomic alterations with a size of 1kb or more [23]. Previous studies have reported that CNV of genes is related to the progression of various cancers through influences on gene expression level [24, 25]. Therefore, we used Pearson correlation analysis to investigate whether there was relationship between gene expression level, mRNAsi, and the CNV of differentially expressed mRNAs.
Establishment of the prognostic risk score model based on prognosis-related genes through Lasso regression

We selected survival-related genes from the differentially expressed miRNAs, IncRNAs, and mRNAs via Univariate Cox PHR. The selected survival-related genes were fitted into Lasso regression to screen candidate genes for establishing a prognostic risk score model by using the “glmnet” package in R. Multivariate Cox PHR was applied to establish the prognostic risk score model based on the regression coefficient of candidate genes through the formula: Risk score= \( \exp(X_1) \times \beta_1 + \exp(X_2) \times \beta_2 + \ldots + \exp(X_n) \times \beta_n \). \( \exp(X) \) refers to the expression of the candidate gene, and \( \beta \) represents the regression coefficient of candidate genes. The accuracy of the model was verified by K-M curve and ROC curve analysis via the “survival” package.

Establishment of a nomogram model based on genes related to prognosis

A nomogram model was built by utilizing the “rms” package in order to more intuitively predict the overall survival of OC at 1, 3, and 5 years based on the prognosis-related genes. The clinical impact and predictive power of the nomogram model were assessed by calibration curves, a DCA curve, and a clinical impact curve.

Analysis of the relationship between mRNAsi and tumor infiltrating immune cells

The relative ratio of 22 tumor-infiltrating immune cell types was calculated following the method of Yue Gao et al, utilizing the CIBERSORT algorithm [26]. The Wilcoxon rank-sum test was employed to distinguish differential infiltration of immune cells between the high- and low-mRNAsi groups. Principal component analysis (PCA) was used to investigate the grouped samples. The relationship of mRNAsi and tumor infiltrating immune cells was explored via Pearson correlation analysis.

Results

Identification of correlations between mRNAsi and clinical characteristics

The mRNAsi score of 302 OC cases were obtained by using the OCLR algorithm, and sample data was divided into two groups based on the median mRNAsi score (Figure 1A). As shown in Figure 1B and 1F, we found no significant correlation between mRNAsi and the age or stage of OC cases, respectively. OC patients with venous or lymphatic invasion had higher mRNAsi scores compared to patients with no invasion (Figure 1C, D). Further, grade 3/4 cases had higher mRNAsi score compared to grade 1/2 cases.
Filtrating of differentially expressed miRNAs, IncRNAs, and mRNAs

By using the “limma” package, 313 lncRNAs (239 downregulated and 74 upregulated), 38 miRNAs (16 downregulated and 22 upregulated), and 2 downregulated mRNAs were obtained and then visualized by a volcano plot (Figure 2A, B, C).

Construction of the IncRNA-miRNA-mRNA competing endogenous RNA networks

The ceRNA network was built based on differentially expressed lncRNA, miRNA, and mRNA using Cytoscape. We first built the interaction network based on gene pair correlations (Figure 2D). We then selected 5 lncRNA-miRNA-mRNA gene pairs to construct the ceRNA network (Figure 2E, Supplementary Table 1). The network contained 5 lncRNAs (ENSG00000230333, ENSG00000227220, ENSG00000234323, ENSG00000233821, ENSG00000266411), 4 miRNAs (hsa-miR-141-5p, hsa-miR-19b-1-5p, hsa-miR-30e-5p, hsa-miR-s29a-5p), and 1 mRNA (CHN2).

Analysis of the relationship between CALB2 and CHN2 copy number variation and mRNAsi score

Figure 3E depicts the locations of CALB2 and CHN2 on the chromosome. Initially, the relationship between the gene expression level and CNV of CALB2 and CHN2 were analyzed. There was a positive correlation between the gene expression level and CNV of CALB2 and CHN2, which was consistent with our assumption (Figure 3A, B). Transcriptomic data indicated that the expression of CALB2 and CHN2 was lower in the high-mRNAsi group than in the low-mRNAsi group, and, based on this, we predicted that the mRNAsi score may have a negative correlation with the CNV of CALB2 and CHN2. Thus, we investigated the relationship between mRNAsi score and the CNV of CALB2 and CHN2. The result indicated that the CNV of CHN2 was negatively correlated with the mRNAsi score (Figure 3C). However, there was no statistically significant correlation between the mRNAsi score and the CNV of CALB2 (Figure 3D).

Establishment of the prognostic risk score model by Lasso regression
Univariate Cox PHR was used to identify survival-related IncRNAs, miRNAs, and mRNAs from differentially expressed genes based on mRNAsi score grouping. In general, 60 IncRNAs and 3 miRNAs (p<0.05) were obtained, as shown in Supplementary Table 2. To further narrow down candidate survival-related genes, 11 candidate genes, including 9 IncRNAs and 2 miRNAs, were obtained by using the Lasso binary logistic regression model (Figure 4A, B). Regression analysis of multivariate Cox proportional hazards was then used to further narrow the scope of candidate genes. 9 candidate genes, including 7 IncRNAs and 2 miRNAs, were obtained and used to establish the risk score model. The risk score model was then built with the addition of expression level results and the regression coefficient of each gene in the Lasso regression model: risk score = 0.00014*ENSG00000230601+0.000213*ENSG00000275850+0.000174*ENSG00000225807 - 0.00052*ENSG00000234754+5.66E-05*ENSG00000259645+4.27E-05*ENSG00000227068 +4.85E-05*ENSG00000249706 -0.02221*hsa-miR-133a-3p+0.099624*hsa-miR-6788-3 (Supplementary Table 3). OC patients with lower risk scores had better OC outcome, as shown in Figure 4C. The ROC curve had an area of under curve of 0.658, which indicated that the prognostic risk score model possessed good predictive power for prognosis (Figure 4D). The risk score of each OC patient was calculated based on expression of the 9 selected genes. The survival status, risk score distribution, and expression pattern of the 9 genes are depicted in Figure 4E-G. I have changed this so that the way you write out Lasso is consistent throughout the text.

**Construction and assessment of a nomogram for overall survival prediction for OC**

A nomogram model based on the 9 candidate genes was established for the prediction of overall survival of OC at 1, 3, and 5 years (Figure 5A). The model’s calibration curve for the probability of overall survival at 1, 3, and 5 years had accurate predictive capacity (Figure 5B-D). A DCA curve was constructed to assess the nomogram model’s clinical benefits. We found that the nomogram model could be of benefit for predicting the 1-, 3- and 5-year overall survival based on the red oblique line of the DCA curve (Figure 5 E-G). Finally, a clinical impact curve was plotted to visualize the nomogram model’s clinical potential. The predicted number of high-risk patients was more than the high-risk number with an adverse event at 1, 3, and 5 years, indicative of the predictive power of the nomogram model (Figure 5 H-J).

**Assessment of the relationship between mRNAsi and tumor-infiltrating immune cells**

We calculated the relative ratio of 22 tumor infiltrating immune cell types by using the CIBERSORT algorithm. We obtained 21 tumor-infiltrating immune cell types, as there was no significant fraction of naïve CD4 T cells. CIBERSORT results are presented in Figure 6A. We then investigated differential immune cell infiltration between the high- and low-mRNAsi groups. As shown in Figure 6B, fractions of follicular T helper cells, M1 macrophages, and activated dendritic cells were much higher in the high-
mRNAsi group compared to the low-mRNAsi group. However, the fraction of resting dendritic cells followed the opposite trend. The other 17 infiltrating immune cell types showed no significant difference. As revealed by PCA analysis results presented in Figure 6C, tumor-infiltrating immune cell fractions could effectively discriminate OC patients belonging to the high-mRNAsi group and low-mRNAsi group. We explored the relationship between mRNAsi score and immune infiltration by Pearson correlation analysis. The result indicated a positive correlation between mRNAsi score and plasma cells, activated dendritic cells, and follicular T helper cells (Figure 7A, E, F). Moreover, mRNAsi decreased in parallel with the increase of resting dendritic cells, as well as M2 macrophages and resting mast cells (Figure 7B, C, D).

Discussion

OC is the most deadliest cancer of the female reproductive system [1], and is usually diagnosed at an advanced stage with poor prognosis and high chance of disease recurrence [5]. CSCs play an important part throughout cancer progression and recurrence, thus also affecting prognosis [9]. However, the exact role of OC stem cells remains unclear. Tumor-infiltrating cells also play crucial roles in the pathogenesis of OC. Whether there is a relationship there between OC stemness and tumor-infiltrating immune cells is currently unclear. IncRNAs play a significant part in the progression of malignant tumors. While studies have reported several IncRNAs with roles in OC pathogenesis, a ceRNA network based on OC stemness has not been previously reported. In order to better understand OC, elucidation of the previously mentioned unclear aspects is urgently needed. In this study, we utilized unsupervised learning in order to gain insights into the relationship between stemness, prognosis, and immune infiltration in OC.

Several studies have reported that mRNAsi could be used for cancer prognosis [27, 28] and the exploration of novel cancer biomarkers [29, 30]. Our results revealed that OC patients with venous or lymphatic invasion had higher mRNAsi scores compared to patients without invasion (Figure 1C, D). In addition, grade 3/4 patients had higher mRNAsi scores compared to grade 1/2 patients (Figure 1E). However, no significant difference in prognosis was observed between the high- and the low-mRNAsi group, which is different from observations in other kinds of cancer [31]. This may due to differences between tumor types or limitations deriving from the current sample size, and requires further investigation. As IncRNAs participate in various aspects of the malignant tumor progression, we screened for differentially expressed IncRNAs, miRNAs, and mRNAs based on the mRNAsi grouping by using the “limma” package. Finally, a ceRNA network containing 5 IncRNAs (ENSG00000230333, ENSG00000227220, ENSG00000234323, ENSG00000233821, ENSG00000266411), 4 miRNAs (hsa-miR-141-5p, hsa-miR-19b-1-5p, hsa-miR-30e-5p, hsa-miR-s29a-5p), and 1 mRNA (CHN2) was constructed by using Cytoscape (Figure2E). We also found that the CNV of the differentially expressed gene CHN2, was negatively correlated with the mRNAsi score (Figure 3C). CHN2 expression, the GTPase-activating protein of p21-rac, is expected to result in higher Rac activity and, therefore, may play a part in the progression from low-grade to high-grade tumors [32, 33]. CALB2, also referred to as Calretinin, is involved in cancer progression, including that of OC. Rathore et al. [34] reported that Calretinin was expressed in 29/31 sex cord stromal tumors, and Yishu Yuan et al. reported that Calretinin and P53, together with PAX8, had a synergistic effect on the differentiation of epithelial OC cells [35]
Univariate Cox PHR was used to identify survival-related IncRNAs, miRNAs, and mRNAs of the differentially expressed genes based on mRNAsi score grouping. We then established a model by Lasso regression, including 7 IncRNAs and 2 miRNAs, for predicting the prognosis of OC patients. The model could efficiently classify OC patients into two groups, a high- and a low-risk group (Figure 4E-D). To further assess whether the 9 candidate genes could be utilized to predict the overall survival rate of OC at 1, 3, and 5 years, we developed a nomogram model. The nomogram model could remarkably predict overall survival of OC at 1, 3, and 5 years. To our best knowledge, such a 9-gene predictive model has not been reported previously. This model may be of use for the identification of novel prognostic markers for OC. However, it is still early to presume that the model is superior to established prognostic examinations. Thus, further experimental validation is necessary. As the nomogram only covered 9 genes, clinical characteristics, including age, lymphatic or/and venous invasion, were not included and may also require further investigation in relation to their prognostic value. Multicenter prospective cohort studies with large sample sizes are required before clinical implementation, as our research was only based on TCGA data with a relatively small sample size.

As various clinical trials of immunotherapy for OC are currently underway [36], we evaluated the potential relationship between mRNAsi and tumor-infiltrating immune cells on the basis of mRNAsi score grouping. Our results showed the fractions of follicular T helper cells, M1 macrophage, and activated dendritic cells were higher in the high-miRNAsi group compared to the low-miRNAsi group. The opposite was true for the fraction of resting dendritic cells. Through Pearson correlation analysis, we found a positive correlation between mRNAsi score and activated dendritic cells, plasma cells, and follicular T helper cells (Figure 7A, E, F, Supplementary Table 4). Analysis also revealed a negative correlation not only between mRNAsi score and resting dendritic and mast cells, but also between mRNAsi score and M2 macrophages (Figure 7B, C, D, Supplementary Table 4). Different tumor infiltrating immune cell types have various roles in cancer, as in the case of cancer-promoting M2 macrophages [37] as well as mast cells which not only enhance proliferation, but also improve the survival of naïve B and activated B cells [38]. This may partially explain why no strong correlation was found between mRNAsi score and OC prognosis.

**Conclusion**

The present study revealed that the mRNAsi score is closely related to immune infiltration in OC. We proposed a prognostic risk model established through Lasso regression, including 7 IncRNAs and 2 miRNAs, which had considerable predictive ability. We also developed a nomogram model, which could remarkably predict overall survival of OC after 1, 3, and 5 years. These models demonstrated potential for clinical application as part of the treatment process for OC.

**List Of Abbreviations**

OC: ovarian cancer; mRNAsi: mRNA expression-stemness index; CSCs: cancer stem cells; TIICs: tumor infiltrating immune cells; OCLR: one-class logistic regression; LncRNA: long non-coding RNAs; ceRNA: competing endogenous RNA; Cox PHR: cox proportional hazard regression model; TCGA: the Cancer
Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files

Competing interests
The authors declare that they have no competing interests

Funding
Not applicable

Authors' contributions
The conception of this article is from Xx M. X P analyzed the data and then wrote the manuscript under the supervision of Xx M.

Acknowledgments
The authors acknowledge the TCGA database for providing high-quality data to researchers. The authors are also grateful for the kind assistance from their colleagues.

References

1. Slatnik CL, Duff E: Ovarian cancer: Ensuring early diagnosis. The Nurse practitioner 2015, 40(9):47-54.

2. Surveillance, Epidemiology, and End Results Program. Cancer stat facts: ovarian cancer. National Cancer Institute (Accessed 2 June 2020), Available at: https://seer.cancer.gov/statfacts/html/ovary.html: National Cancer Institute.
3. Jessmon P, Boulanger T, Zhou W, Patwardhan P: Epidemiology and treatment patterns of epithelial ovarian cancer. Expert review of anticancer therapy 2017, 17(5):427-437.

4. Chien J, Poole EM: Ovarian Cancer Prevention, Screening, and Early Detection: Report From the 11th Biennial Ovarian Cancer Research Symposium. International journal of gynecological cancer : official journal of the International Gynecological Cancer Society 2017, 27(9S Suppl 5):S20-s22.

5. Motohara T, Masuda K, Morotti M, Zheng Y, El-Sahhar S, Chong KY, Wietek N, Alsaadi A, Karaminejadranjbar M, Hu Z et al: An evolving story of the metastatic voyage of ovarian cancer cells: cellular and molecular orchestration of the adipose-rich metastatic microenvironment. Oncogene 2019, 38(16):2885-2898.

6. Bast RC, Jr., Hennessy B, Mills GB: The biology of ovarian cancer: new opportunities for translation. Nature reviews Cancer 2009, 9(6):415-428.

7. Vargas AN: Natural history of ovarian cancer. Ecancermedicalscience 2014, 8:465.

8. Nash Z, Menon U: Ovarian cancer screening: Current status and future directions. Best practice & research Clinical obstetrics & gynaecology 2020, 65:32-45.

9. Chen W, Dong J, Haiech J, Kilhoffer MC, Zeniou M: Cancer Stem Cell Quiescence and Plasticity as Major Challenges in Cancer Therapy. Stem cells international 2016, 2016:1740936.

10. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, Minden M, Paterson B, Caligiuri MA, Dick JE: A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 1994, 367(6464):645-648.

11. Bonnet D, Dick JE: Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature medicine 1997, 3(7):730-737.

12. Visvader JE, Lindeman GJ: Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nature reviews Cancer 2008, 8(10):755-768.

13. Ahmed N, Abubaker K, Findlay JK: Ovarian cancer stem cells: Molecular concepts and relevance as therapeutic targets. Molecular aspects of medicine 2014, 39:110-125.

14. Bapat SA, Mali AM, Koppikar CB, Kurrey NK: Stem and progenitor-like cells contribute to the aggressive behavior of human epithelial ovarian cancer. Cancer research 2005, 65(8):3025-3029.

15. Scarlett UK, Rutkowski MR, Rauwerdink AM, Fields J, Escovar-Fadul X, Baird J, Cubillos-Ruiz JR, Jacobs AC, Gonzalez JL, Weaver J et al: Ovarian cancer progression is controlled by phenotypic changes in dendritic cells. The Journal of experimental medicine 2012, 209(3):495-506.

16. Hamanishi J, Mandai M, Iwasaki M, Okazaki T, Tanaka Y, Yamaguchi K, Higuchi T, Yagi H, Takakura K, Minato N et al: Programmed cell death 1 ligand 1 and tumor-infiltrating CD8+ T lymphocytes are prognostic factors of human ovarian cancer. Proceedings of the National Academy of Sciences of the United States of America 2007, 104(9):3360-3365.

17. Obermeyer Z, Emanuel EJ: Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016, 375(13):1216-1219.
18. Sokolov A, Paull EO, Stuart JM: **ONE-CLASS DETECTION OF CELL STATES IN TUMOR SUBTYPES.** Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing 2016, 21:405-416.

19. Chen P, Fang X, Xia B, Zhao Y, Li Q, Wu X: **Long noncoding RNA LINC00152 promotes cell proliferation through competitively binding endogenous miR-125b with MCL-1 by regulating mitochondrial apoptosis pathways in ovarian cancer.** Cancer medicine 2018, 7(9):4530-4541.

20. Cheng FHC, Lin HY, Hwang TW, Chen YC, Huang RL, Chang CB, Yang W, Lin RI, Lin CW, Chen GCW et al: **E2F6 functions as a competing endogenous RNA, and transcriptional repressor, to promote ovarian cancer stemness.** Cancer science 2019, 110(3):1085-1095.

21. Liang H, Yu T, Han Y, Jiang H, Wang C, You T, Zhao X, Shan H, Yang R, Yang L et al: **LncRNA PTAR promotes EMT and invasion-metastasis in serous ovarian cancer by competitively binding miR-101-3p to regulate ZEB1 expression.** Molecular cancer 2018, 17(1):119.

22. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O et al: **Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation.** Cell 2018, 173(2):338-354.e315.

23. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, Xu D, Fan X: **Copy number variation is highly correlated with differential gene expression: a pan-cancer study.** BMC medical genetics 2019, 20(1):175.

24. Yang L, Wang YZ, Zhu HH, Chang Y, Li LD, Chen WM, Long LY, Zhang YH, Liu YR, Lu J et al: **PRAME Gene Copy Number Variation Is Related to Its Expression in Multiple Myeloma.** DNA and cell biology 2017, 36(12):1099-1107.

25. Gut A, Moch H, Choschzick M: **SOX2 Gene Amplification and Overexpression is Linked to HPV-positive Vulvar Carcinomas.** International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists 2018, 37(1):68-73.

26. Gao Y, Chen L, Cai G, Xiong X, Wu Y, Ma D, Li SC, Gao Q: **Heterogeneity of immune microenvironment in ovarian cancer and its clinical significance: a retrospective study.** Oncoimmunology 2020, 9(1):1760067.

27. Pei J, Wang Y, Li Y: **Identification of key genes controlling breast cancer stem cell characteristics via stemness indices analysis.** Journal of translational medicine 2020, 18(1):74.

28. Xia P, Li Q, Wu G, Huang Y: **Identification of Glioma Cancer Stem Cell Characteristics Based on Weighted Gene Prognosis Module Co-Expression Network Analysis of Transcriptome Data Stemness Indices.** Journal of molecular neuroscience : MN 2020.

29. Zhao M, Chen Z, Zheng Y, Liang J, Hu Z, Bian Y, Jiang T, Li M, Zhan C, Feng M et al: **Identification of cancer stem cell-related biomarkers in lung adenocarcinoma by stemness index and weighted correlation network analysis.** Journal of cancer research and clinical oncology 2020, 146(6):1463-1472.

30. Pan S, Zhan Y, Chen X, Wu B, Liu B: **Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices.** Frontiers in oncology 2019, 9:613.
31. Zhang Y, Tseng JT, Lien IC, Li F, Wu W, Li H: mRNAsi Index: Machine Learning in Mining Lung Adenocarcinoma Stem Cell Biomarkers. *Genes* 2020, **11**(3).

32. Casado-Medrano V, Barrio-Real L, García-Rostán G, Baumann M, Rocks O, Caloca MJ: A new role of the Rac-GAP β2-chimaerin in cell adhesion reveals opposite functions in breast cancer initiation and tumor progression. *Oncotarget* 2016, **7**(19):28301-28319.

33. Elaimy AL, Guru S, Chang C, Ou J, Amante JJ, Zhu LJ, Goel HL, Mercurio AM: VEGF-neuropilin-2 signaling promotes stem-like traits in breast cancer cells by TAZ-mediated repression of the Rac GAP β2-chimaerin. *Science signaling* 2018, **11**(528).

34. Rathore R, Arora D, Agarwal S, Sharma S: Correlation of FOXL2 with Inhibin and Calretinin in the Diagnosis of Ovarian Sex Cord Stromal Tumors. *Turk patoloji dergisi* 2017, **33**(2):121-128.

35. Yuan Y, Zhu W, Gong G, Jiang J, Wang Y, Wu X: Molecular markers derived from ovarian cancer and its correlation with the types of oviductal epithelial lesions. *Journal of Central South University Medical sciences* 2019, **44**(5):528-534.

36. Wang W, Liu JR, Zou W: Immunotherapy in Ovarian Cancer. *Surgical oncology clinics of North America* 2019, **28**(3):447-464.

37. Cheng H, Wang Z, Fu L, Xu T: Macrophage Polarization in the Development and Progression of Ovarian Cancers: An Overview. *Frontiers in oncology* 2019, **9**:421.

38. Merluzzi S, Betto E, Ceccaroni AA, Magris R, Giunta M, Mion F: Mast cells, basophils and B cell connection network. *Molecular immunology* 2015, **63**(1):94-103.