

Development of a Prognostic Model for Colorectal Cancer Based on Four EMT-related Genes

Shuoyang Hung

Department of Gastrointestinal Surgery, Wuhan University Renmin Hospital <https://orcid.org/0000-0003-4870-3816>

Chao Yang

Department of Gastrointestinal Surgery, Renmin Hospital of Wuhan University

Fengyu Cao

Department of Gastrointestinal Surgery, Renmin Hospital of Wuhan University

Xiaobo He

Department of Gastrointestinal Surgery, Renmin Hospital of Wuhan University

Yongbin Zheng (✉ yongbinzheng@whu.edu.cn)

Department of Gastrointestinal Surgery, Renmin Hospital of Wuhan University

Research

Keywords: EMT, colorectal cancer, signature, TCGA, GEO, nomogram

DOI: <https://doi.org/10.21203/rs.3.rs-135628/v1>

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

[Read Full License](#)

Abstract

Background: Epithelial-to-mesenchymal transition (EMT) greatly affects the progression and metastasis of colorectal cancer (CRC) which is one of the most malignant tumors. The goal of this study was to evaluate the prognostic value of EMT-related genes and develop a prognostic nomogram for patients' overall survival (OS) with CRC.

Methods: This retrospective study analyzed gene expression profiles and clinical data in 1099 samples of CRC patients from the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Risk signature based on public databases was created and examined for its predictive value in CRC patients. The independence and correlation of risk signature with clinicopathological parameters were respectively evaluated through univariate/multivariate analysis and subgroup analysis. A nomogram was established according to risk signature and clinical parameters and validated for its predictive ability. The correlation of risk signature and infiltrating immune cells in CRC was assessed.

Results: A four EMT-related genes (ERGs) based signature was constructed and validated for its ability in prognostic prediction in CRC patients. The results showed the ERGs signature as an independent prognostic factor regarding different clinicopathological parameters including age, gender and clinical stage. Besides, the ERGs signature also performed well in subgroup analysis with T3-T4, N-, M0 and early disease stage. A nomogram based on the ERGs signature combining clinical parameters was constructed which showed a promising predictive ability in prognosis of CRC patients' OS. In addition, we found a positive correlation of the ERGs signature with six different types of immune cells.

Conclusion: A novel ERGs signature was identified in this study and ERGs signature can be useful in predicting the prognosis of CRC patients.

1. Introduction

Colorectal cancer (CRC) is the most common cause of gastrointestinal disease related death, ranking third of modality in malignant tumors(1). With a high metastasis-recurrence rate, CRC presents a dismal prognosis for most patients. According to the latest global cancer statistics in 2018, approximately 1.8 million people are diagnosed with CRC and 900,000 patients die from the disease each year(2). Patients with CRC typically undergo surgical treatment, and advancing surgical methods have contributed to ameliorated overall prognosis(3–5). Besides, with the affiliation of adjuvant chemoradiotherapy the overall survival and relapse-free survival for CRC patients have been improved profoundly(3, 6). Additionally, neoadjuvant chemotherapy can increase the percentage of patients who are eligible for resection although the improvement of OS from neoadjuvant chemotherapy remains controversial(7–9). However, uncontrolled metastasis and recurrence still occur in CRC patients despite radical curative surgery and chemoradiotherapy and account for the main cause for the death of CRC patients(10) and the current approaches of assessing CRC patients' prognosis are not precise enough.

Thus, more effective prediction of prognosis for CRC patients are urgently needed for better clinical management.

Epithelial-to-mesenchymal transition (EMT) is regarded as a crucial regulatory process in mediating invasion and metastasis in cancer. EMT enables epithelial cells to acquire a series of mesenchymal characteristics, like motility and invasiveness(11). And recently, studies have shown that epithelial cells are allowed to gain mesenchymal phenotype in which case cells transiently exhibit both epithelial and mesenchymal features without completely transforming into mesenchymal cells as assumed by the classical definition of EMT(12). In contrast with EMT, mesenchymal-epithelial transition (MET) provides cancer cells the loss of mesenchymal phenotypes and the ability of relocating into a distant metastasis. Increasing evidence presents the importance of controlling the EMT process for preventing progression and metastasis in multiple cancers, including CRC, and methods on treatment of targeting EMT process are also in researching(13). However, the clinicopathological character of EMT in CRC has not been well understood. Therefore, we performed systemic bioinformatic analysis in this study to explore the potential role of EMT in prediction of patient prognosis and discover the possible novel modality targeting EMT process to improve existing treatment.

In this study, we filtrated four EMT-related genes (ERGs) from differentially expressed genes (DEGs) by analyzing gene datasets in Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) data sets. We evaluated the correlation of ERGs signature with clinical characteristics and tumor associated immune cells through which the risk signature may reflect CRC microenvironment and immune response. Moreover, we integrated the ERGs signature with clinical factors to build a prognostic nomogram, which allowed ameliorated prognosis assessment of CRC patients.

2. Methods

2.1 Study design and data collection

Study design of this research is presented in the flow chart (Fig. 1A). Level 3 RNA sequencing data (RNA-seq) and corresponding clinical information of 514 CRC patients was obtained from the TCGA database (<https://portal.gdc.cancer.gov/repository>)(14). The gene expression profiles were normalized by scale method in "limma" R package. RNA-seq data and paired clinical information of another 585 CRC patients was downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>)(15) and corresponding dataset was GSE39582. Normalized count values were used. Both TCGA and GEO database are publicly available, thus the present study was exempted from the approval of local ethics committees. Criteria for study inclusion were: (1) Repeated tumor samples in the same patient were removed. (2) Patients with unknown survival status and follow-up information, and those who died within a follow-up period of 30 days were excluded. (3) Patients with unknown disease stage or grade were excluded. Consequently, 416 patients meeting the criteria were included in the discovery set and 532 patients in validation set. The TCGA discovery set was used to establish an ERG signature to predict patients' prognosis, while the GEO set were used for validation.

200 ERGs were downloaded from the Gene Set Enrichment Analysis (GSEA) database (<http://software.broadinstitute.org/gsea/index.jsp>)(16).

2.2 Screening of differentially expressed ERGs

Differentially expressed mRNA between tumor and adjacent normal tissue samples were screened in TCGA cohort by the “limma” package of R software 4.0.2. $|\log_2 \text{ fold change}| > 1$ and false discovery rate (FDR) < 0.05 were set as the cut-off criteria. DEGs in paired samples from TCGA datasets, characterized by upregulation and downregulation, were intersected with GSEA ERGs set to obtain differentially expressed ERGs.

2.3 Construction and validation of an ERG signature

After identifying hub ERGs, we performed univariate Cox analysis of overall survival (OS) to screen ERGs with prognostic value. Subsequently, we performed least absolute shrinkage and selection operator (LASSO) Cox regression analysis in TCGA discovery set to select the best gene model for predicting prognosis in CRC patients. LASSO Cox regression analysis was performed using R package “glmnet” and the optimal values of penalty parameters were determined by 10-fold cross-validation. Furthermore, Kaplan-Meier (K-M) survival curves were used to assess survival difference between high- and low-risk groups using the “survminer” package in R software. The risk score model for each patient was determined by a linear combination of gene expression weighted by the regression coefficient from LASSO Cox regression analysis. Then, we used the median value of the risk score as cut-off value which was next used to divide patients into high- and low-risk group. To assess the predictive accuracy of the ERG signature time-dependent Receiver Operator characteristic (ROC) analysis was performed with area under the curve (AUC) calculated at different cut-off time using the “survivalROC” package in R software.

2.4 Prognostic value validation of the ERGs signature in an independent data set

The same ERGs signature model was used in GEO set for external validation. The risk score was calculated for each patient in validation set and X-tile software(17) was utilized for determining the optimum cut-off point. Survival analysis was performed in GEO cohort to validate the prognostic ability of ERG-signature. Additionally, time-dependent ROC curves were employed to evaluate the prediction accuracy of ERG-signature model in validation set.

2.5 Independence of the prognostic signature from other clinicopathological parameters

To determine the independence of the prognostic model from other clinical parameters, univariate and multivariate analysis of risk-score with age, gender, disease stage was performed. The relationship between the risk score and clinicopathological parameters was showed in forest plots.

2.6 Subgroup analysis of correlation between the ERG signature and other clinicopathological parameters

To evaluate whether the prognostic model was associated with other clinicopathological parameters including age, gender, disease stage, pathological T stage, pathological N stage and pathological M stage, patients were divided into subgroups of age ≤ 65 years and age ≥ 65 years, female and male, pathological tumor stage I-II and stage III-IV, pathological T1-T2 and T3-4, pathological N⁻ and N⁺ (N1-3), without metastasis M0 and with metastasis M1. K-M survival analysis was performed for the subgroups of different clinicopathological parameters.

2.7 Nomogram construction and validation

A nomogram integrating the ERG signature and various clinicopathological factors was established for predicting patients' prognosis using the "rms" package in R. Calibration curves were applied to assess the accuracy of the nomogram.

2.8 Estimation of relative abundance of immune cell types

The fraction data of six immune cells, including B cell, CD4⁺ T cell, CD8⁺ T cell, Neutrophil, Macrophage and Dendritic, was downloaded from the Tumor Immune Estimation Resource (TIMER) website (<https://cistrome.shinyapps.io/timer/>)(18) and their correlation with risk score was calculated.

2.9 Statistical analysis

The heat map was generated using the "pheatmap" package in R. Student's t-test was used to compare subgroups and paired data. Pearson's chi-square test was applied to analyze differences between the discovery and validation sets and the association between the risk score and clinical parameters. K-M survival curves were compared using log-rank test. Univariate Cox regression was conducted to estimate the hazard ratio (HR) for different factors. Multivariate Cox regression was performed to determine independent factors. All statistical analyses were conducted by R software. $P < 0.05$ was considered significant.

3. Result

3.1 Identification of differentially expressed ERGs in CRC

Gene expression profiles of CRC patients from TCGA database were extracted and intersected with 200 EMT-related genes downloaded from GSEA dataset. Then, we compared gene expression levels of the filtrated genes between tumor tissues and corresponding normal tissues and identified 62 DEGs of which 23 were down-regulated and 39 were up-regulated. According to the DEGs a heatmap was drawn and the distribution of DEGs expression was displayed in the Volcano plot (Fig. 1B&C).

3.2 Identification of a four-ERG signature in the TCGA discovery set

We used Univariate Cox analysis to identify ERGs with prognostic value and seven ERGs with highest Hazard ratios were screened out and a forest plot was performed (Fig. 1D). To identify ERGs with most

valuable prognostic efficacy, we performed the least shrinkage and selection operator (LASSO) Cox regression model in the TCGA discovery set. Consequently, we obtained four ERGs, namely PCOLCE2, OXTR, BDNF, SERPINE1 to construct our gene signature. Besides, all the four ERGs were negatively correlated with patient survival.

Based on the respective expression levels, we constructed a risk score formula: Risk score = $0.246592982 \times \text{Exp}(\text{BDNF}) + 0.070329939 \times \text{Exp}(\text{OXTR}) + 0.052615049 \times \text{Exp}(\text{SERPINE1}) + 0.276502574 \times \text{Exp}(\text{PCOLCE2})$. (Exp: expression level of the gene)

Expression levels of the four ERGs in TCGA cohorts were exhibited in a heatmap (Fig. 2A). Patients were separated into high- and low-risk score groups according to a median cut-off value of 0.546. The distribution of risk scores and survival time were shown in Fig. 2B&2C. Poorer prognosis was presented in the high-risk group than the low-risk one. A time-dependent ROC analysis was performed to show the prognostic accuracy of the four-gene signature. The AUC was 0.639 at 1 year, 0.664 at 3 year and 0.603 at 5 year (Fig. 2D). K-M survival analysis corroborated that patients in the high-risk group showed a poorer overall survival rate than the low-risk ($p < 0.01$) (Fig. 2E).

3.3 Validation of four-ERG signature

To further verify the prognostic efficacy of four-ERG signature, we examined the prognostic potential in a GEO validation set (Fig. 3A). Risk-score of each patient was calculated by the same formula but the patients in GEO cohort was separated into high- and low-risk groups using an optimum cut-off point (2.6) determined by X-tile software (Fig. 3B). Consistent with the result in discovery set, high-risk patients in validation set had a poorer prognosis than low-risk group (Fig. 3C). Besides, ROC curve performed well in the validation set with AUC of 0.615 at 1 year, 0.583 at 3 year and 0.574 at 5 year (Fig. 3D). Similarly, a lower survival rate was exhibited in high-risk patients from validation set by survival analysis (Fig. 3E).

3.4 Association with clinicopathological factors and subgroup analysis

To verify four-ERG signature was an independent prognostic factor regarding clinical characteristics, we performed univariate and multivariate cox analysis (Fig. 4). The result showed that the risk score could predict the prognosis independently regarding to age, gender and clinical stage.

To further evaluate the prognostic value of four-ERG signature in specific clinical conditions, subgroup analysis between risk score and different clinical parameters was performed. The result suggested that risk signature stayed powerful in either younger or older, male or female. However, the result was quite intriguing in subgroups of tumor node metastasis (TNM) stage and disease stage. Our risk signature showed a convincing predictive power in subgroup of T3-T4 ($p = 0.001$), N⁻ ($p < 0.001$), M0 ($p < 0.001$) and early disease stage (stage I-II, $p < 0.001$) while high-risk group also tended to have a poorer prognosis but the difference didn't reach significance in T1-2 ($p = 0.282$, N⁺ ($p = 0.213$) and advanced disease stage (stage III-IV, $p = 0.290$). Besides, the risk score showed no efficient predictive ability in the metastasis (M1) subgroup ($p = 0.714$) (Fig. 5).

3.5 Construction and validation of ERG-based nomogram

With the aim to develop a feasible tool for clinical predicting OS of CRC patients, we constructed a nomogram with the risk score (based on the ERGs signature), age, gender and disease stage in the TCGA dataset (Fig. 6A). Calibration plots showed agreeable concordance between observed and predicted rates of 1-, 3- and 5-year OS. Besides, time-dependent ROC curves were used for assessing the prognostic predictive ability of the nomogram and other predictors. The result indicated that the nomogram owned satisfying accuracy of 1-, 3- and 5-year OS in combination with age, gender and disease stage (Fig. 6B).

3.6 The correlation of the four-ERG signature with immune cells infiltration.

To assess the efficacy of ERGs signature reflecting the tumor immune microenvironment status, we evaluated the risk score and the representation of different immune cells. The result showed that ERGs signature was positively associated with CD4⁺ T cells (Cor = 0.459, $p < 0.01$), CD8⁺ T cells (Cor = 0.247, $p < 0.01$), B cells (Cor = 0.223, $p < 0.01$), dendritic cells (Cor = 0.453, $p < 0.01$), macrophages (Cor = 0.513, $p < 0.01$) and neutrophils (Cor = 0.379, $p < 0.01$) (Fig. 7). These data suggested that the four-ERG signature could serve as an indicator of immune status of CRC.

4. Discussion

In consideration of the importance of epithelial-mesenchymal transition in malignancy metastasis(19), developing meaningful gene signatures to monitor the EMT status and motility of tumor cells in cancer patients is significant not only to identify reliable prognostic biomarkers but, if applied properly, to stratify patients at high-risk of recurrence, metastasis or chemoresistance who might benefit from additional therapy. In the present study, we validated a prognostic signature, based on four ERGs, which proved to be a reliable indicator of EMT status and could identify CRC patients with unfavorable prognosis. Moreover, our prognostic signature can further stratify patients sharing specific clinicopathological factors (e.g., age, gender and disease stage) into subgroups with different survival outcomes. Integrating these findings, we constructed a nomogram, incorporating the ERGs signature's risk score and clinical characteristics, which showed good performance for predicting survival in patients with CRC.

Although several gene signatures representative of tumor EMT status have been proposed with potential clinical applicability in several cancers(16, 20), the impact of EMT status in CRC progression and prognosis prediction remains to be fully explored. Subgroup analysis combining our ERGs signature risk score and clinicopathological parameters revealed that a higher risk score was more strongly correlated with poorer prognosis. This is owing to high risk score reflecting the possibilities of enhancing the motility and invasiveness of tumor cells contributing to tumor progression and metastasis. Noteworthy, we also confirmed that the ERG signature remained an independent prognostic factor with multivariate analysis after adjusting for clinicopathological variables.

Cancer progression and metastasis is often linked to altered plasticity of tumor cells(21, 22), while EMT plays an unignorable role in this process(23). Accordingly, aberrant expression of transcription factors, cytokines, epithelial surface markers and mesenchymal features has been shown to contribute to the migration of CRC cells(24). According to previous studies, the four ERGs in our novel CRC signature encode proteins with active participation in angiogenesis, developing resistance and chemotaxis, increasing cancer cell growth, proliferation, survival, migration and epithelial to mesenchymal transition, and decreasing anoikis, relapse and chemotherapeutic sensitivity.

Brain-derived neurotrophic factor (BDNF) is highly expressed in brain as member of neurotrophins which were initially known to function as growth factors in nervous system(25). However, recent studies have discovered that BDNF and its receptor tropomyosin receptor kinase B (TrkB) were upregulated in a wide range of cancers including CRC(26). Growing evidence showed that BDNF/TrkB pathway was closely involved in CRC proliferation, invasion, migration and poor prognosis(27–29). Besides, positive results render BDNF/TrkB pathway to be a promising therapeutic target for CRC treatment(27, 30).

Oxytocin (OT) is viewed as a hypothalamic neuropeptide and classically functions in uterine contraction and milk ejection(31). Interestingly, some findings suggested that the gene encoding OT receptor (OXTR) also expressed in several types of cancer cells including breast, ovarian and prostate cancer regulating proliferation(32, 33). The relation between OXTR and CRC hasn't been validated though it was reported that oxytocin receptors were expressed throughout gastrointestinal tract(34).

Plasminogen activator inhibitor 1 & 2 (PAI-1 & PAI-2), also known as serpin family E member 1 & 2 (SERPINE1 & SERPINB2), are the main inhibitor of urokinase plasminogen activator (uPA) and tissue plasminogen activator (tPA) which constitute plasminogen activator (PA) system. PA system in turn acts crucially in extracellular matrix (ECM) remodeling and further plays an important role in tumor progression(35). However, SERPINE1 was found associating with tumor progression and metastasis and identified as an indicator of poor prognosis in a series of cancers(36–38). Similarly, the character of SERPINE1 (or PAI-1) in colorectal cancer was also identified, as a prognostic biomarker(39, 40).

PCOLCE2 encodes procollagen C-proteinase enhancer 2 which is collagen-binding protein capable of binding at multiple sites on the triple helical portions of fibrillar collagens and competing for such binding with procollagen C-proteinases(41). In consistent with our research, PCOLCE2 has already been identified as a diagnostic and prognostic biomarker in CRC patients(42). However, the underlying mechanism of PCOLCE2 engaging in the progression and metastasis of CRC remains to be elucidated and we for the first time propose the idea that PCOLCE2 is associated with EMT process of CRC cancer cells.

Accumulating evidence shows that tumor cells undergoing EMT or acquiring mesenchymal phenotype might have contribution to the component of tumor infiltrating immune cells (TIICs) shifting to an immunosuppressive pattern resulting in immune escape and immune resistance(43, 44). Studies have also showed cross-talk between EMT and different immune cells render the availability to predict the presentation of TIICs by assessing EMT status(20, 45–47). In this study, ERG signature showed a positive

correlation with immune cells including CD4⁺ T cells, dendritic cells and macrophages. In the light of our knowledge, CD4⁺ T cells and macrophages contribute greatly to the formation of immunosuppressive microenvironment, however, different subpopulations of these immune cells exhibit diverse functions and prognostic roles in CRC, thus immune cells can't be simply regarded as risk or protective factors to CRC(46). Nevertheless, our study showed a certain change in infiltrating immune cells regarding to the ERGs, indicating the potential value of further investigating the relationship between the ERGs and subpopulations of different tumor-associated immune cells and their prognostic and therapeutic role in CRC.

5. Conclusions

In summary, we identified a gene signature for CRC based on four differentially expressed ERGs, which could act as an indicator for tumor EMT status and to classify patients into groups with distinctly different prognosis. Additionally, we constructed a composite prognostic nomogram for prediction of CRC patients' prognosis integrating ERGs signature and clinical factors, which may serve as a promising clinical tool. Limitation of our study lay in its retrospective nature and the need to define the molecular mechanisms by which the ERGs in our signature impact CRC progression and metastasis. Nevertheless, besides aiding current prognostic efforts, we hope the data provided will lead to further studies aimed at exploring novel therapeutic strategies for CRC.

Abbreviations

CRC: colorectal cancer

DEGs: differentially expressed genes

EMT: epithelial-to-mesenchymal transition

ERGs: EMT-related genes

OS: overall survival

TCGA: The Cancer Genome Atlas

GEO: Gene Expression Omnibus

MET: mesenchymal-epithelial transition

GSEA: Gene Set Enrichment Analysis

FDR: false discovery rate

LASSO: least absolute shrinkage and selection operator

K-M: Kaplan-Meier

ROC: time-dependent Receiver Operator characteristic

AUC: area under the curve

HR: hazard ratio

Cor: correlation coefficient

BDNF: brain-derived neurotrophic factor

TrkB: tropomyosin receptor kinase B

OT: oxytocin

OXTR: OT receptor

PAI-1 & PAI-2: plasminogen activator inhibitor 1 & plasminogen activator inhibitor 2

SERPINE1 & SERPINB2: serpin family E member 1 & serpin family E member 2

uPA: urokinase plasminogen activator

ECM: extracellular matrix

PCOLCE2: procollagen C-proteinase enhancer 2

TIIcS: tumor infiltrating immune cells

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 81372553)

Authors' contributions

SG and CY made substantial contributions to conception and design of the research. SH, CY and FC carried out data collection, analysis. SH, CY and FC provided interpretation of the results. SH and CY wrote the paper. SH, CY, FC, XH, and YZ edited the manuscript and provided critical comments. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
2. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol.* 2019;16(12):713-32.
3. Punt CJ, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat Rev Clin Oncol.* 2017;14(4):235-46.
4. Akgul O, Cetinkaya E, Ersoz S, Tez M. Role of surgery in colorectal cancer liver metastases. *World J Gastroenterol.* 2014;20(20):6113-22.
5. De Rosa M, Pace U, Rega D, Costabile V, Duraturo F, Izzo P, et al. Genetics, diagnosis and management of colorectal cancer (Review). *Oncol Rep.* 2015;34(3):1087-96.
6. van Steenbergen LN, Elferink MAG, Krijnen P, Lemmens V, Siesling S, Rutten HJT, et al. Improved survival of colon cancer due to improved treatment and detection: a nationwide population-based study in The Netherlands 1989-2006. *Ann Oncol.* 2010;21(11):2206-12.
7. Khan K, Wale A, Brown G, Chau I. Colorectal cancer with liver metastases: neoadjuvant chemotherapy, surgical resection first or palliation alone? *World J Gastroenterol.* 2014;20(35):12391-406.
8. Nordlinger B, Sorbye H, Glimelius B, Poston GJ, Schlag PM, Rougier P, et al. Perioperative chemotherapy with FOLFOX4 and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC Intergroup trial 40983): a randomised controlled trial. *The Lancet.* 2008;371(9617):1007-16.

9. Nordlinger B, Van Cutsem E, Rougier P, Kohne CH, Ychou M, Sobrero A, et al. Does chemotherapy prior to liver resection increase the potential for cure in patients with metastatic colorectal cancer? A report from the European Colorectal Metastases Treatment Group. *Eur J Cancer*. 2007;43(14):2037-45.
10. Al Bandar MH, Kim NK. Current status and future perspectives on treatment of liver metastasis in colorectal cancer (Review). *Oncol Rep*. 2017;37(5):2553-64.
11. Ye X, Weinberg RA. Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. *Trends Cell Biol*. 2015;25(11):675-86.
12. Bakir B, Chiarella AM, Pitarresi JR, Rustgi AK. EMT, MET, Plasticity, and Tumor Metastasis. *Trends Cell Biol*. 2020.
13. Singh M, Yelle N, Venugopal C, Singh SK. EMT: Mechanisms and therapeutic implications. *Pharmacol Ther*. 2018;182:80-94.
14. The Cancer Genome Atlas. <https://portal.gdc.cancer.gov/repository>. Accessed on 20 Sep 2020.
15. Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/>. Accessed on 20 Sep 2020.
16. Cao R, Yuan L, Ma B, Wang G, Qiu W, Tian Y. An EMT-related gene signature for the prognosis of human bladder cancer. *J Cell Mol Med*. 2020;24(1):605-17.
17. Camp, R. L. JCCCR. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. 2004;10(21):7252-9.
18. Tumor Immune Estimation Resource. <https://cistrome.shinyapps.io/timer/>. Accessed on 23 Sep 2020.
19. Mittal V. Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu Rev Pathol*. 2018;13:395-412.
20. Thompson JC, Hwang WT, Davis C, Deshpande C, Jeffries S, Rajpurohit Y, et al. Gene signatures of tumor inflammation and epithelial-to-mesenchymal transition (EMT) predict responses to immune checkpoint blockade in lung cancer with high accuracy. *Lung Cancer*. 2020;139:1-8.
21. Te Boekhorst V, Friedl P. Plasticity of Cancer Cell Invasion-Mechanisms and Implications for Therapy. *Adv Cancer Res*. 2016;132:209-64.
22. Friedl P, Alexander S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell*. 2011;147(5):992-1009.
23. Lu W, Kang Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Dev Cell*. 2019;49(3):361-74.
24. Vu T, Datta PK. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. *Cancers (Basel)*. 2017;9(12).
25. Meng L, Liu B, Ji R, Jiang X, Yan X, Xin Y. Targeting the BDNF/TrkB pathway for the treatment of tumors. *Oncol Lett*. 2019;17(2):2031-9.
26. Meldolesi J. Neurotrophin Trk Receptors: New Targets for Cancer Therapy. *Rev Physiol Biochem Pharmacol*. 2018;174:67-79.
27. Mazouffre C, Geyl S, Perraud A, Blondy S, Jauberteau MO, Mathonnet M, et al. Dual inhibition of BDNF/TrkB and autophagy: a promising therapeutic approach for colorectal cancer. *J Cell Mol Med*.

- 2017;21(10):2610-22.
28. Fujikawa H, Tanaka K, Toiyama Y, Saigusa S, Inoue Y, Uchida K, et al. High TrkB expression levels are associated with poor prognosis and EMT induction in colorectal cancer cells. *J Gastroenterol.* 2012;47(7):775-84.
 29. Yu Y, Zhang S, Wang X, Yang Z, Ou G. Overexpression of TrkB promotes the progression of colon cancer. *APMIS.* 2010;118(3):188-95.
 30. Akil H, Perraud A, Jauberteau MO, Mathonnet M. Tropomyosin-related kinase B/brain derived-neurotrophic factor signaling pathway as a potential therapeutic target for colorectal cancer. *World J Gastroenterol.* 2016;22(2):490-500.
 31. Zingg HH, Laporte SA. The oxytocin receptor. *Trends in Endocrinology & Metabolism.* 2003;14(5):222-7.
 32. Lerman B, Harricharran T, Ogunwobi OO. Oxytocin and cancer: An emerging link. *World J Clin Oncol.* 2018;9(5):74-82.
 33. Cassoni P, Sapino A, Marrocco T, Chini B, Bussolati GJJJoN. Oxytocin and oxytocin receptors in cancer cells and proliferation. 2010;16(4):362-4.
 34. Monstein HJ, Grahn N, Truedsson M, Ohlsson B. Oxytocin and oxytocin-receptor mRNA expression in the human gastrointestinal tract: a polymerase chain reaction study. *Regul Pept.* 2004;119(1-2):39-44.
 35. Pavón MA, Arroyo-Solera I, Céspedes MV, Casanova I, Oncotarget RMJ. uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy. 2016;7(35).
 36. McMahon BJ, Kwaan HC. Components of the Plasminogen-Plasmin System as Biologic Markers for Cancer. *Adv Exp Med Biol.* 2015;867:145-56.
 37. Seker F, Cingoz A, Sur-Erdem I, Erguder N, Erkent A, Uyulur F, et al. Identification of SERPINE1 as a Regulator of Glioblastoma Cell Dispersal with Transcriptome Profiling. *Cancers (Basel).* 2019;11(11).
 38. Liao P, Li W, Liu R, Teer JK, Xu B, Zhang W, et al. Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer. *Onco Targets Ther.* 2018;11:6969-80.
 39. Wang Q, Lu W, Yin T, Lu L. Calycosin suppresses TGF-beta-induced epithelial-to-mesenchymal transition and migration by upregulating BATF2 to target PAI-1 via the Wnt and PI3K/Akt signaling pathways in colorectal cancer cells. *J Exp Clin Cancer Res.* 2019;38(1):240.
 40. Dong W, Wu X. Overexpression of Rab11-FIP2 in colorectal cancer cells promotes tumor migration and angiogenesis through increasing secretion of PAI-1. *Cancer Cell Int.* 2018;18:35.
 41. Steiglitz BM, Keene DR, Greenspan DS. PCOLCE2 encodes a functional procollagen C-proteinase enhancer (PCPE2) that is a collagen-binding protein differing in distribution of expression and post-translational modification from the previously described PCPE1. *J Biol Chem.* 2002;277(51):49820-30.

42. Chen L, Lu D, Sun K, Xu Y, Hu P, Li X, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene*. 2019;692:119-25.
43. Terry S, Savagner P, Ortiz-Cuaran S, Mahjoubi L, Saintigny P, Thiery JP, et al. New insights into the role of EMT in tumor immune escape. *Mol Oncol*. 2017;11(7):824-46.
44. Jiang Y, Zhan H. Communication between EMT and PD-L1 signaling: New insights into tumor immune evasion. *Cancer Lett*. 2020;468:72-81.
45. Wei C, Yang C, Wang S, Shi D, Zhang C, Lin X, et al. Crosstalk between cancer cells and tumor associated macrophages is required for mesenchymal circulating tumor cell-mediated colorectal cancer metastasis. *Molecular Cancer*. 2019;18(1).
46. Li S, Xu F, Zhang J, Wang L, Zheng Y, Wu X, et al. Tumor-associated macrophages remodeling EMT and predicting survival in colorectal carcinoma. *Oncol Immunology*. 2017;7(2).
47. Chae YK, Chang S, Ko T, Anker J, Agte S, Iams W, et al. Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci Rep*. 2018;8(1):2918.