Metastasis progression through the interplay between the immune system and Epithelial-Mesenchymal-Transition in circulating breast tumor cells

Samane Khoshbakht
University of Tehran https://orcid.org/0000-0003-3253-7577
Sadegh Azimzadeh Jamalkandi (azimzadeh.jam.sadegh@gmail.com)
University of Tehran https://orcid.org/0000-0003-3403-3700
Ali Masudi-Nejad
University of Tehran

Primary research

Keywords: breast cancer, single CTC, cluster CTC, metastasis, co-expression, EMT, immune response

Posted Date: June 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-32145/v1

Abstract

Background

Circulating tumor cells (CTCs) are the critical initiator of systemic dissemination of cancer, contributing to distant metastasis formation. The metastatic cascades rely on the fundamental roles of different types of CTCs. In which the dual immune responses and epithelial-mesenchymal-transition (EMT) are of two metastasis-driving phenomena and require more molecular assessments.

Methods

In this study, we investigated the transcriptomic modular pattern of single/cluster circulating tumor cells (CTCs). The co-expression analysis implemented, and we could detect two metastatic subnetworks indicating the immune responses and EMT in CTCs. Furthermore, a directed subnetwork identified in the KEGG database. The metastatic potential of subnetworks assessed and validated by classification methods on primary tumors. And, we could fit risk models to distant-metastasis survival of patients.

Results

Our results show the crosstalk among EMT, immune system, menstrual cycles, and stemness in CTCs. In which, fluctuation of menstrual cycles (hormone-related signals) is a new detected pathway in CTCs in breast cancer. The immune SVM model showed high metastatic potential in classifying patients metastatic/non-metastatic groups (accuracy, sensitivity, and specificity scores are 78%). The distant-metastasis free survival model could be used to stratify patients into low, medium, and high-risk groups. Finally, PTCRA, F13A1, LAT, ICAM2, and SNRPC are novel detected biomarkers in breast cancer.

Conclusion

In conclusion, different types of CTCs, including cluster/single cells, are metastasis-leading elements in breast cancer. In which, individual assessment of their intrinsic biological properties may assist elucidating metastasis-related mechanisms. These findings may apply to develop superior treatments in the clinic.

Background

Metastasis is the common cancer-associated cause of deaths that accounts for a remarkable 90% deaths [1]. Cancer progression and metastasis are of the critical and even controversial aspects of cancer biology. There are two arguable metastasis models, including parallel progression and linear progression, which try to explain the dark side of the cancer metastasis[2]. In the linear model, the tumor initiates by genetic and epigenetic alternations, grow, spread, and gain metastasis potentials and disseminate to ectopic sites. Still, in the parallel model, the metastasis ability initiates early-onset and separately evolve [2, 3]. Apart from multiple metastasis models and molecular mechanisms which indicate a different aspect of cancer initiation and progression, the circulating tumor cells (CTCs) in patients' bloodstream

and likewise their physical characteristics of single CTC or clustered CTC play a crucial role in metastasis propensity [4]. Of note, CTCs are rare disseminated tumor cells in the peripheral blood of patients that are negatively related to the high rise of mortality rates in cancer. They borrow the morphologic features of their primary tumor and gain new features to survive and metastasize secondary organs [5]. The cluster CTCs consist of 2-50 cancer cells within the circulation of patients, and they have a 23- to 50-fold increase in metastasis potential[4].

Circulating tumor cells overcome many hurdles to colonize distant organs [6] and includes intravasation into blood or lymph, evading immune bulwarks, extravasation to distant sites, and eventually replacing with host tissue microenvironment [6, 7]. The CTCs mirror their primary tumor characteristics and even metastases they initiate, with which strong abilities to survive under shear forces in circulation and eventually overtake host tissue [6, 7].

The two important power of malignant cells is the reversal phenotypic of Epithelial-mesenchymal transition (EMT) and even immune system suppression. EMT is a cellular transition in which cells acquire mesenchymal characteristics that can accelerate metastasis through immunosuppression [8, 9]. The epithelial cells that isolated the primary site undergo immediately anoikis, a fate likely to meet most CTCs in blood circulation. Accordingly, such mesenchymal transformation may provide longer survival signals to reduce apoptotic outcome [4, 10].

Metastasis is the leading cause of death among women with breast cancer, and CTCs are prominent and leading components that appear even in early stages [11, 12]. Detection of CTCs in metastatic and non-metastatic breast cancer patients implies its role in cancer progression. Therefore, fully realize the CTCs' molecular characteristics will guide us to unknown metastasis concepts and more precise therapeutic decisions.

In this study, we implemented the co-expression network reconstruction for CTCs isolated from advanced patients. We extracted metastasis relevant subnetworks that enriched in the immune system and EMT. The metastasis-free survival of genes assessed in GSE7390. Concerning a better understanding of signaling inside CTCs, we also extracted an induced directed subnetwork from the KEGG database. We also carried out the support vector machine (SVM) classification for selected subnetwork on GSE7390 to prepare a predictive model to classify metastatic and non-metastatic primary tumors. The subnetworks preservations were validated on another dataset.

Materials And Methods

2.1 | data sets and metadata information

The single-cell RNA-seq data related to advanced ER+ breast cancer patients were downloaded from the NCBI data repository (GSE86978). The data consist of 77 cells in which 47 of them are clustered CTCs, 22 CTCs are single cells, and the rest of the cells are not categorized. The GSE51827 was used for subnetwork preservation analysis that consists of 29 cells (single CTCs and cluster CTCs), in which 15

cells are single CTCs, and 14 cells are cluster CTCs. The GSE7390 consists of 198 breast cancer patients' expression extracted from the primary tumors. The patients are not treated. The GSE9195 consists of 77 breast cancer treated with tamoxifen, which is used for classification and metastasis-free survival validation.

2.2 | Pre-processing, normalization, and differential analysis

To have more precise downstream analysis and remove non-biological variations, we implemented several pre-process steps on genes and also cells. At the first step, we filtered out low abundance genes, then the small count cells omitted subsequently. In the last step, the expression data were normalized to reduce technical effects using the scatter package [13]. The differentially expressed analysis (DEA) was completed using the limma package in R [14]. The DEA was implemented to compare clustered cells expression to the single cells' expression (FDR < 0.05).

2.3 | Co-expression network reconstruction (CNNR) and subnetwork extraction

The co-expression network reconstructed using a weighted correlation network analysis (WGCNA) method [15]. The pairwise relation among genes was estimated using the Pearson correlation among genes. Concerning having more connected subnetworks, we carried out the topological overlap matrix (TOM) and, consequently, connectivity gene filtering (connectivity values less 0.1 are omitted). Eventually, we used hierarchical clustering to extract subnetworks. The trait used in this study is clustered and single status of the cells captured in blood. The subnetworks in which have strong correlations between their first principle component and the biological trait are selected as trait related subnetworks. The gene significance and module membership were used to filter out essential genes in selected subnetworks. The gene significance is the correlation between gene expression and the trait. The module membership is the correlation between gene expression and module representative (first principle component in the principal component analysis (PCA)).

2.4 | Directed network reconstruction

We downloaded all homo sapiens pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database resource [16]. All the KEGG Mark-up Language (KGML) pathways parsed and converted to graph objects, and finally, these graphs merged [17]. Furthermore, the KEGG ids annotated to gene symbols. At the last step, we extracted KEGG induced subnetwork using the most trait associated subnetworks (correlation > 0.5) resulted from the CNNR step. The genes were categorized by their biological process feature using ClueGO plug-in in Cytoscape [18, 19]. The network visualization implemented by the Cytoscape and the Gephi software [19, 20].

2.5 | Gene set enrichment analysis and subnetwork preservation analysis

The most trait-related subnetworks ( ) extracted from CNNR, were enriched using ConsensusPathDB webserver (q-value < 0.05) [21]. The hierarchical clustering applied to selected subnetworks to check the genes' potential to separate cluster CTCs and single CTCs.

The GSE51827 downloaded from NCBI to implement preservation analysis of subnetworks in another dataset in R. The data pre-processed, normalized, and merged using scatter package, which is suitable for single-cell RNA-seq data. The preservation combined statistics including and were used to check the reproducibility of subnetworks [22]. These two combined statistics consist of 12 statistics, which all of them are calculated.

The includes connectivity and density statistics, which shows the interaction pattern among genes in subnetworks. The subnetworks with are not preserved. if , the subnetwork is semi preserved, and if , the subnetwork is preserved. Also, a higher indicates more preservation of subnetworks in the second data set [22].

2.6 | Distant metastasis classification

To evaluate the importance of selected subnetworks and the potential of genes in metastasis, we implemented the classification algorithms on two individual datasets. The expression data refer to primary tumors. The GSE7390 (Affymetrix platform, HG-U133A) downloaded using the GEOquerry package in R [23]. The ER+ patients (134 patients out of 198 ones) filtered, and the expression data normalized using the RMA method [24]. In this section, we learned three classifiers, including support vector machine (SVM), artificial neural network (ANN), and decision tree on metastatic and non-metastatic patients [25]. The classification algorithms ran with and without feature selection algorithms, including the genetic algorithm (GA) and the world competitive contest (WCC) algorithm. The SVM ran with 5-fold cross-validation and 80 percent of cells as the training set. Finally, the accuracy, precision, and specificity were checked to select a better classifier for metastasis prediction, furthermore to extract the most metastatic features too. The selected model was assessed in another dataset (GSE9195).

2.7 | Distant metastasis-free survival analysis

The Kaplan-Meire distant metastasis-free survival estimate and overall survival analysis were implemented using GSE7390 in R[26]. The patients were stratified due to quantiles. The expression values lower than the second quantile were labeled low expression, and expression values higher than the fourth quantile were labeled high expression. The stepwise Cox proportional hazard ratio (Cox-PH) modeling was implemented on selected subnetworks [26]. The Variance Inflation Factor (VIF) lower than two was used as the variable selection criteria. The second and fourth quantiles of the predicted hazard ratio were used for stratifying patients into three groups, including low-risk, medium-risk, and high-risk patients.

Results

3.1 | Pre-processing of CTCs and DEA

There are 74 out of 77 cells after pre-processing for downstream analyses. The excluded cells had low quality; therefore, we omited them. The differential expression analysis was implemented after the

normalization step. The expression difference between cluster CTCs and single CTCs groups were assessed (FDR < 0.05). The genes of immune subnetwork are downregulated (light purple color in the heatmap) in cluster CTCs compare to single CTCs. Furthermore, the genes of the EMT subnetworks imply upregulation (dark purple color in the heatmap) in cluster CTCs (Fig. 1). The immune and EMT subnetworks were used independently to cluster all cells. As illustrated in Fig. 1, the single and cluster CTC cells separated well with selected subnetworks. The immune-related subnetwork represents a stronger expression difference between cluster cells and single cells.

3.2 | Metastasis associated subnetworks

Metastasis associated subnetworks were determined by co-expression analysis and hierarchical clustering. We detected 16 subnetworks. The first principle component (in PCA analysis) of subnetworks and the trait (cluster CTCs vs. single CTCs) relationship assessed by correlation analysis. The two top subnetworks with the highest correlation with the trait, nominated for Enrichment (midnightblue correlation =0.57, turquoise correlation = 0.51). The midnightblue and turquoise subnetworks sizes are 35 and 22 genes, respectively (Fig. S1). The midnightblue subnetwork enriched for immune responses and the turquoise subnetwork enriched for EMT (Fig. 2).

The EMT subnetwork enriched for cancer related pathways such as cell-cell communication, tight junction, keratinization, estrogen signaling pathway. The immune subnetwork enriched for pathways such as platete activation, immune system, innate immune system.

To have a biological concept for subnetworks, we address the midnightblue and the turquoise subnetworks, the immune and the EMT subnetworks, respectively. The immune-related novel genes are PTCRA, F13A1, LAT, GNG11, ICAM2, NRGN, P2RX1, CLEC1B, BIN2, LPAR5, CCL5, SELP, RUFY1, C6orf25, TUBB1, GFI1B, C2orf88, ACRBP, andC17orf72. The list of genes identified in the immune-related subnetwork is Table S1.

The EMT related genes are LRPPRC, AGR2, CLDN4, CRIP1, DSP, ELF3, JUP, KRT8, KRT18, KRT19, FAM102A, TACSTD2, EPCAM, PEBP1, PSMD8, RAN, SNRPC, SPTAN1, EZR, DDR1, MLPH, and WDR34. In which, SNRPC is a metastatic novel gene in breast cancer that is upregulated in CTC clusters. The list of genes identified in EMT-related subnetwork is summarized in Table S2.

The preservation of all subnetworks was assessed in another dataset (GSE51827). The two combined statistics and calculated to assess subnetworks preservation in the second data set (immune subnetwork: , and EMT subnetwork: ,). The values for both subnetworks are more than 10, and values are high too (min= 7, max= 32). Subsequently, these statistics represent our immune and EMT subnetworks preservation in the second data set (Table S3).

3.2 | Directed network reconstruction

The signaling pathways inside the CTCs are not well recognized. The 336 homo sapiens signaling pathways out of 537 ones in the KEGG database downloaded and merged. The association among two

selected subnetworks (immune and EMT) were investigated by extracting induced subnetwork from KEGG. A directed subnetwork of size 255 genes extracted and illustrated in Fig. 3. There are 12 gene categories obtained from biological processes, including Hormonal regulation, Immune responses, Ion metabolism, Nucleobase metabolism, Oxidative responses, Protein localization, Protein topology response, STAT singling pathway, Vitamin metabolism, cell differentiation, circulation in blood regulation, Energy metabolism. These categories are illustrated by colors on network nodes, and the genes with no category remained grey. The nodes with multiple colors were detected in different biological processes. The node size is illustrated by the node degrees. PLCG1 and ENTPD8 are two hub nodes in the network that participate in energy and nucleobase metabolisms, respectively.

3.3 | Distant metastasis classification model

The distant metastasis potential of two nominated subnetworks for classifying patients into primary tumors was assessed using SVM, neural network, and decision tree methods. The 5-fold cross-validation SVM accuracy, sensitivity, and specificity scores for EMT related subnetwork are 79%, 78%, and 21%, respectively. The neural network accuracy, sensitivity, and specificity scores are 18%, 18%, and 80%, respectively. Eventually, the decision tree accuracy, sensitivity, and specificity scores are 60%, 60%, and 30%, respectively. These results refer to a full model (all genes in the subnetwork included). Compare to these three models, the SVM model is the strongest method in classifying metastatic and non-metastatic patients, but the specificity score is too low. The SVM model validated in GSE9195.

The SVM accuracy, sensitivity, and specificity scores for immune-related subnetwork are 78%, 78%, and 78%, respectively. The neural network accuracy, sensitivity, and specificity scores are 85%, 85%, and 14%, respectively. Eventually, the decision tree accuracy, sensitivity, and specificity scores are 71%, 71%, and 36%, respectively. These results refer to a full model (all subnetwork genes included). The specificity of the neural network and decision tree method is low compare to the SVM. Due to results, the SVM model is the most powerful method in classifying metastatic and non-metastatic patients for the immune-related subnetwork. The immune-related subnetwork accuracy, sensitivity, and specificity for the SVM model are superior to EMT related subnetwork.

The feature selection methods were implemented. The accuracy, sensitivity, and specificity of feature selection models (GA and WCC) did not improve. The feature selection algorithms implemented in the SVM model for both subnetworks. The WCC introduced 13 genes, and GA introduced 12 genes for EMT related subnetwork. The WCC algorithm introduced 15 genes, including HLA-E, MYLK, WIPF1, TLN1, F13A1, NRGN, ICAM2, PTGS1, SELP, PF4, ITGA2B, GFI1B, TUBB1, PTCRA, RUFY1, BIN2, and CLEC1B for EMT subnetwork. The GA introduced 17 genes for immune-related subnetworks, including CCL5, MYLK, WIPF1, TLN1, NRGN, GNG11, PTGS1, SELP, ITGA2B, MAX, GFI1B, P2RX1, PTCRA, RUFY1, and BIN2. The SVM model (full model) for immune subnetwork validated in GSE9195. The accuracy, sensitivity, and specificity are 0.868, which is superior to GSE7390. The results confirm that the immune-related genes detected in this study can classify metastatic and non-metastatic samples more precisely compared to the neural network and decision tree models in two data sets. We implemented the classification methods

to assess the metastasis potential of two nominated subnetworks which extracted from CTCs' data. We fitted the Cox-PH as a predictive model to discriminate patients to low, medium, and high metastasis risk groups.

3.4 | Distant metastasis and overall survival analysis

The association between gene expression and distant metastasis-free survival /overall survival was implemented to detect metastasis potential genes in selected subnetworks. The JUP, KRT18, and KRT19 overall survival and distant metastasis-free survival are significant by log-rank test (p-value < 0.05) (Fig. 4a, b, c, d, e, and f). These three genes belong to EMT subnetwork. The high expression level of JUP, KRT18, and KRT19 associate with low overall survival and further lower metastasis progression. The distant metastasis-free survival and overall survival curves indicate the same pattern.

We fitted a metastasis free-survival Cox-PH regression model for EMT and Immune subnetworks to assess metastasis progression and patients' predictive models. The Immune Cox-PH model includes of RUFY1 and P2RX1 variables (Likelihood ratio test p-value = 0.0295). The EMT Cox-PH model includes of RAN, PEBP1, KRT8, DSP, DDR1, and CLDN4 variables (Likelihood ratio test p-value = 0.0001016). The variables' coefficients and p-values reported in Table S4 and S5. All the significant genes in the model have VIF < 2 to avoid multicollinearity (Table S6 and S7). The proportional hazard assumption for two model variables was assessed by the Schoenfeld residuals (Fig. S2 and S3). The predictive Cox-PH models for distant metastasis-free survival for two subnetworks are illustrated in Fig. 4g and h. The concordance index, as a model evaluation measure, for EMT and Immune predictive Cox-PH models are 0.7 and 0.6, respectively. Therefore, the EMT model is more powerful in discriminating patients into low, medium, and high metastasis risk groups compare to the Immune model (higher concordance index indicates more power in discrimination).

Discussions

Whereas multiple studies on circulating tumor cells (CTCs) as single CTCs or metastatic microemboli (CTC clusters) have been conducted, the molecular mechanisms of such rare cells are insufficiently characterized. Several metastatic potentials of CTC to overcome many restrictions on extravasation of the primary tumor microenvironment, survival in the bloodstream, and successfully colonize secondary organs are still a mystery. Therefore, a better understanding of the molecular features of CTCs' types is needed.

This study aims to explore metastasis clues. We have implemented the co-expression analysis to detect subnetworks discriminating single/cluster CTCs (Fig. 1). Two of subnetworks indicated a high correlation to the trait (single/cluster status of CTCs). These two illustrate immune- and EMT-related pathways that mediate bloodborne dissemination of cancer cells (Figs. 2a and b). Due to previous studies, the immune-associated mechanisms and EMT are of two major arms in breast cancer progression and metastasis, but investigating them in CTCs is not studied well [27]. To prepare cancer cells for intravasation, the keratin family, claudins, and cadherins must be downregulated through the EMT process in primary

tumors. But, due to the surviving urgency of CTCs in the bloodstream, and also avoiding anoikis, a small number of tumor cells must be attached and break off from the primary site [28, 29]. Therefore, the keratins, claudins, and cadherins should be upregulated in CTC clusters to survive shear forces in blood circulation. The plakoglobin (JUP), KRT8, KRT18, KRT19, CLDN4, and EPCAM are such essential genes that their role in breast cancer metastasis demonstrated in previous studies [4, 29, 30].

The KRT8, KRT18, KRT19 are a group of cytoskeleton genes within the cellular cytoplasm called keratins. Although they extensively used as diagnostic tumor markers, several studies have demonstrated their involvement in cancer cell invasion and metastasis, as well as in treatment responsiveness. Keratins are the intermediate filament-forming proteins of epithelial cells that organize the internal three-dimensional cellular structure also act in cell shape maintenance by bearing tension. [31]. Therefore, they may play an essential role in CTC clusters in the bloodstream shear forces. The plakoglobin is one of the cell junction genes that hold tumor cells together in CTC clusters. And its upregulation in breast cancers CTC clusters compare to single cells demonstrated in Aceto, Nicola, et al. study [4]. In our study, the overexpression of JUP, KRT18, and KRT19, as well as the overall survival and metastasis-free survival, are significant either (Fig. 4a or d). The EPCAM and cytokeratins are reported as detection markers in the Enrichment of CTCs [32]. These markers guide scientists to detect metastatic patients. But, the rest of the genes are not investigated in CTC studies, including CTC detection or molecular mechanisms. So, they may be a good alternative in cluster CTCs studies. There are several types of immune cells that ambiguously reveal anti- and pro-tumor behaviors. The immunosuppressive microenvironment of tumors protects the primary tumor cells. But, while tumor cells extravasate and enter circulation, they lose their tumor protection. Therefore, they must adapt themselves to escape immune surveillance [1, 33]. The interplay between immune cells and cytokeratins contributes to CTCs evasion from immune surveillance. The cytotoxic T lymphocytes (CTLs) recruited by recognizing tumor antigens presented by major histocompatibility class I (MHCI) [34, 35]. The under-expression of MHCI in tumor cell surface guides them to hide from CTLs, and thereby survive in circulation. Moreover, the overexpression of cytokeratins such as KRT8, and together with heterodimeric partners KRT18 and KRT19 inhibit MHCI interactions with CTLS [33, 34]. All these foundings are consistent with our results (overexpression of KRT8, KRT18, and KRT19; under-expression of HLA-E) for CTC clusters, which show the CTC cluster potential to evade the immune system (Fig. 1a and b). Several studies support the association between EMT and immune cell escape [36, 37]. Moreover, a plethora of genes and signals support stem-cell support pathways such as Wnt, TGF-β, and NOTCH [6]. Downregulation of DAB2 (putative tumor suppressor) is reported in breast cancer, which promotes EMT and also involves in the TGF-β pathway [38, 39]. DAB2 downregulation might be related to the stemness phenotype acquired from EMT, which helps CTC clusters to escape the immune system. DAB2 is founded in our immune subnetwork, and it is downregulated in CTC clusters (Fig. 1b). Of note, although the CTC clusters have higher metastatic potential due to less frequency in metastatic patients, the single cells contribute metastasis either. Several studies such as Szczerba, Barbara Maria, et al. indicate more single-cell detection in metastatic patients (about 88.0%) [40]. Therefore, biological mechanisms of metastasis in the single and clusters CTCs need to be investigated in more detail.

The coordination among different cancer-associated pathways is reported in various studies. Atashgaran, Vahid, et al. investigated the crosstalk between the immune response pathway and hormonal regulation pathways, such as the fluctuation of menstrual cycles. The dis-regulation of hormonal factors may affect on genome instability and decrease of immune surveillance [41]. Accordingly, such pro-metastatic crosstalks in breast cancer contribute to progression and metastasis. The crosstalk among immune responses, EMT, and stemness pathways such as embryonic pathways are reported in Takebe, et al. study. The cancer stemness signaling pathway includes ErbB, Wnt, and transforming growth factor (TGF)-β pathways are essential during embryogenesis, moreover, in the stimulation of self-renewing in tumor cells [42]. Such pro-metastatic pathways and the interplay among all of them may happen in circulating tumor cells (Fig. 2c and Fig. 3). These pro-metastatic cells may show different signaling pathways and interrelationship among them to act like multi-role players with great metastatic potentials. The CTCs are tumor cells that show primary tumor characteristics and likewise more additional metastatic propensity to survive in blood-stream and extravasation capabilities. As a result, characterizing multiple aspects of CTCs' involvement in cancer progression is essential and useful in patients' treatment strategies.

Conclusions

Circulating tumor cells as single CTCs or CTC clusters indicate metastatic potentials in ER+ breast cancer patients. However, the CTC clusters show much higher pro-metastatic potentials. They benefit several metastatic capabilities to extravasate, surviving in blood-stream, and finally extravasate to secondary tissues.

CTCs include considerable dis-regulated pathways such as hormonal pathways, EMT, embryonic pathways, and also immune responses. Identification of several cancer-associated pathways in CTCs offers a continuum of potential therapeutic targets. Exclusively, the fluctuation of menstrual cycles and their effect on ER+ breast cancer pathways are less considered in the clinic. The interplay among all pathways and key genes that mediate such bridges among metastatic pathways are essential in cancer biology study and accordingly in therapeutic guidelines.

Abbreviations

CTC: circulating tumor cell

EMT: epithelial-mesenchymal-transition

SVM: support vector machine

PTCRA: Pre T Cell Antigen Receptor Alpha

F13A1: Coagulation Factor XIII A Chain

LAT: Linker For Activation Of T Cells

ICAM2: Intercellular Adhesion Molecule 2

OS: overall survival

PCA: principle component analysis

WCC: world competitive contest

GA: genetic algorithm

ANN: artificial neural network

Cox-PH: Cox proportional hazard ratio

DEG: differential expressed gene

VIF: Variance Inflation Factor

FDR: false discovery rate

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The public data used in this article. The GSE numbers included in the article.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors' contributions

Study concept and design: Khoshbakht and Azimzadeh. Data analysis and interpretation: Khoshbakht, Azimzadeh. Manuscript drafting: Khoshbakht. Statistical analysis: Khoshbakht. Study supervision: Masudi-Nejad and Azimzadeh.

Acknowledgments

Not applicable

References

1. Leone, K., C. Poggiana, and R.J.D. Zamarchi, *The interplay between circulating tumor cells and the immune system: from immune escape to cancer immunotherapy.* Diagnostics, 2018. 8(3): p. 59.
2. Klein, C.A.J.N.R.C., *Parallel progression of primary tumours and metastases.* Nat Rev Cancer, 2009. 9(4): p. 302-312.
3. Ghajar, C.M. and M.J.J.N. Bissell, *Metastasis: pathways of parallel progression.* Nature, 2016. 540(7634): p. 528-529.
4. Aceto, N., et al., *Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis*. Cell, 2014. 158(5): p. 1110-1122.
5. Yang, C., et al., *Circulating tumor cells in precision oncology: clinical applications in liquid biopsy and 3D organoid model.* Cancer Cell International, 2019. 19(1): p. 341.
6. Massagué, J. and A.C.J.N. Obenauf, *Metastatic colonization by circulating tumour cells.* Nature, 2016. 529(7586): p. 298-306.
7. Dasgupta, A., A.R. Lim, and C.M.J.M.o. Ghajar, *Circulating and disseminated tumor cells: harbingers or initiators of metastasis?* Mol Oncol, 2017. 11(1): p. 40-61.
8. Kudo-Saito, C., et al., *Cancer metastasis is accelerated through immunosuppression during Snail-induced EMT of cancer cells.* Cancer cell, 2009. 15(3): p. 195-206.
9. Pastushenko, I. and C.J.T.i.c.b. Blanpain, *EMT transition states during tumor progression and metastasis.* Trends Cell Biol, 2019. 29(3): p. 212-226.
10. Mani, S.A., et al., *The epithelial-mesenchymal transition generates cells with properties of stem cells.* Cell, 2008. 133(4): p. 704-715.
11. Lang, J.E., et al., *RNA-Seq of circulating tumor cells in stage II–III breast cancer.* Ann Surg Oncol, 2018. 25(8): p. 2261-2270.
12. Weigelt, B., J.L. Peterse, and L.J.J.N.r.c. Van't Veer, *Breast cancer metastasis: markers and models.* Nat Rev Cancer, 2005. 5(8): p. 591-602.
13. McCarthy, D.J., et al., *Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.* Bioinformatics, 2017. 33(8): p. 1179-1186.
14. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and computational biology solutions using R and Bioconductor*. 2005, Springer. p. 397-420.

15. Langfelder, P. and S.J.B.b. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. 9(1): p. 559.
16. Zhang, J.D. and S. Wiemann, *KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor.* Bioinformatics, 2009. 25(11): p. 1470-1.
17. Csardi, G. and T.J.I. Nepusz, complex systems, *The igraph software package for complex network research.* InterJournal, complex systems, 2006. 1695(5): p. 1-9.
18. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.* Bioinformatics, 2009. 25(8): p. 1091-1093.
19. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. 13(11): p. 2498-2504.
20. Bastian, M., S. Heymann, and M. Jacomy. *Gephi: an open source software for exploring and manipulating networks*. in *Third international AAAI conference on weblogs and social media*. 2009.
21. Kamburov, A., et al., *ConsensusPathDB: toward a more complete picture of cell biology.* Nucleic Acids Res, 2011. 39(suppl_1): p. D712-D717.
22. Langfelder, P., et al., *Is my network module preserved and reproducible?* PLoS computational biology, 2011. 7(1).
23. Davis, S. and P.S.J.B. Meltzer, *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.* Bioinformatics, 2007. 23(14): p. 1846-1847.
24. Gautier, L., et al., *affy—analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. 20(3): p. 307-315.
25. Masoudi-Sobhanzadeh, Y., H. Motieghader, and A.J.B.b. Masoudi-Nejad, *FeatureSelect: a software for feature selection based on machine learning approaches.* BMC Bioinformatics, 2019. 20(1): p. 170.
26. Therneau, T., *A Package for Survival Analysis in S. version 2.38*. 2015.
27. Santisteban, M., et al., *Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells.* Cancer Res, 2009. 69(7): p. 2887-2895.
28. Fabisiewicz, A. and E.J.M.O. Grzybowska, *CTC clusters in cancer progression and metastasis.* Med Oncol, 2017. 34(1): p. 12.
29. Joosse, S.A., et al., *Changes in keratin expression during metastatic progression of breast cancer: impact on the detection of circulating tumor cells.* Clin Cancer Res, 2012. 18(4): p. 993-1003.
30. Tőkés, A.-M., et al., *Claudin-1,-3 and-4 proteins and mRNA expression in benign and malignant breast lesions: a research study.* Breast Cancer Res, 2005. 7(2): p. R296.
31. Karantza, V.J.O., *Keratins in health and cancer: more than mere epithelial cell markers.* Oncogene2011. 30(2): p. 127-138.
32. Deng, G., et al., *Enrichment with anti-cytokeratin alone or combined with anti-EpCAM antibodies significantly increases the sensitivity for circulating tumor cell detection in metastatic breast cancer patients.* Breast Cancer Research, 2008. 10(4): p. R69.

33. Mohme, M., S. Riethdorf, and K.J.N.r.C.o. Pantel, *Circulating and disseminated tumour cells—mechanisms of immune surveillance and escape.* Nat Rev Clin Oncol, 2017. 14(3): p. 155.
34. Wu, M.-S., et al., *Cytokeratin 8-MHC class I interactions: a potential novel immune escape phenotype by a lymph node metastatic carcinoma cell line.* Biochem Biophys Res Commun, 2013. 441(3): p. 618-623.
35. Joosten, S.A., L.C. Sullivan, and T.H.J.J.o.i.r. Ottenhoff, *Characteristics of HLA-E restricted T-cell responses and their role in infectious diseases.* Journal of immunology research, 2016. 2016.
36. Jia, D., et al., *Quantifying cancer epithelial-mesenchymal plasticity and its association with stemness and immune response.* J Clin Med, 2019. 8(5): p. 725.
37. Terry, S., et al., *New insights into the role of EMT in tumor immune escape.* Molecular oncology, 2017. 11(7): p. 824-846.
38. Bagadi, S.A.R., et al., *Frequent loss of Dab2 protein and infrequent promoter hypermethylation in breast cancer.* Breast Cancer Res Treat, 2007. 104(3): p. 277-286.
39. Martin, J., B.-S. Herbert, and B.J.B.j.o.c. Hocevar, *Disabled-2 downregulation promotes epithelial-to-mesenchymal transition.* Br J Cancer, 2010. 103(11): p. 1716-1723.
40. Szczerba, B.M., et al., *Neutrophils escort circulating tumour cells to enable cell cycle progression.* Nature, 2019. 566(7745): p. 553-557.
41. Atashgaran, V., et al., *Dissecting the biology of menstrual cycle-associated breast cancer risk.* Frontiers in oncology, 2016. 6: p. 267.
42. Takebe, N., R.Q. Warren, and S.P. Ivy, *Breast cancer growth and metastasis: interplay between cancer stem cells, embryonic signaling pathways and epithelial-to-mesenchymal transition.* Breast cancer research, 2011. 13(3): p. 211.

Figures

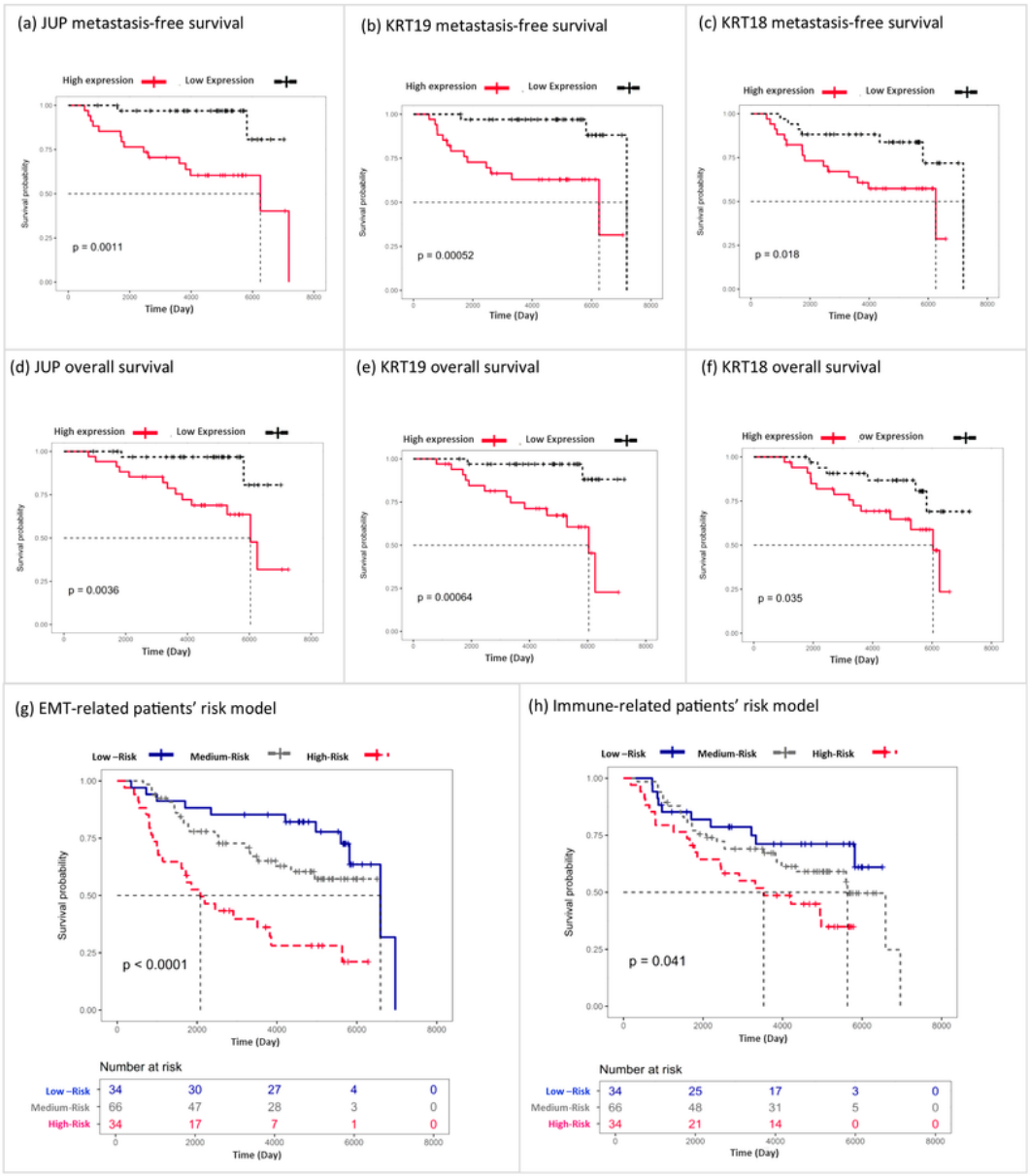

Figure 4

Metastasis free survival and overall survival. p indicate p-value in a,b,c,d,e,f,g, and h section. (g) a predictive risk model for EMT subnetwork. Low risk indicates upper-quartile; low-risk indicates lower-quartile. (h) a predictive risk model for the immune subnetwork. Low risk indicates upper-quartile; low-risk indicates lower-quartile.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.docx