Identification of Potential Tumor-educated Platelets RNA Biomarkers for Pan Cancer Detection and Staging

Xinxin Ge
Soochow University
Liuxia Yuan
Soochow University
Ying Hu
Soochow University
Bin Cheng
Soochow University
Khan Muhammad Shoaib
Soochow University
Kesheng Dai (kdai@suda.edu.cn)
Soochow University

Research article

Keywords: tumor educated platelets, diagnosis, platelet mRNA, bioinformatics analysis

Posted Date: August 26th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-60440/v1

Identification of potential tumor-educated platelets RNA biomarkers for pan cancer detection and staging

Xinxin Ge,[1] Liuxia Yuan,[1] Ying Hu,[1] Bin Cheng,[1] Khan Muhammad Shoaib, [1] Kesheng Dai[1]*

[1] Soochow University, Jiangsu Institute of Hematology, Cyrus Tang Medical Institute, The First Affiliated Hospital and Collaborative Innovation Center of Hematology, State Key Laboratory of Radiation Medicine and Protection, Key Laboratory of Thrombosis and Hemostasis, Ministry of Health, National Clinical Research Center for Hematological Diseases, Suzhou, 215006;

*Correspondence**:** Kesheng Dai.

Jiangsu Institute of Hematology, Cyrus Tang Medical Institute, The First Affiliated Hospital and Collaborative Innovation Center of Hematology, Soochow University, Key Laboratory of Thrombosis and Hemostasis, Ministry of Health, National Clinical Research Center for Hematological Diseases, Suzhou 215006, China.

Tel/Fax: 86 512 67781370.

E-mail: kdai@suda.edu.cn.

Abstract

Backgroud and purpose: Cancer is a disturbing disease with high morbidity and mortality. Although medical technology has been rapidly developed, diagnosis of cancer is still complicated, difficult, as well as expensive. Platelets mRNA profiles are altered after educated by tumors. Thereby tumor-educated platelets (TEPs) mRNA profiles have the potential diagnostic value for cancers. The aim of this study was to identify diagnostic tumor educated platelets (TEPs) genes involved in early or advanced pan cancer.

Patients and methods: 285 platelets samples covering 55 healthy samples, 39 early cancer samples and 191 advanced cancer samples, as well as 636 samples covering 234 healthy samples, 55 early cancer samples and 345 metastatic cancer samples were retrieved from the GEO database (GSE68086 and GSE89843, respectively). The TEPs differentially expressed genes (TEPs DEGs) between healthy samples and early/advanced cancer samples were obtained. Gene ontology (GO) analysis and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis were used to identify the pathways and functional annotation of TEPs DEGs. Protein–protein interaction of these TEPs DEGs was analyzed based on the STRING database and visualized by Cytoscape software. In addition, correlation analysis and diagnostic analysis were performed to evaluate the diagnostic value of TEPs mRNAs expression for early/advanced pan cancer.

Results and conclusions: By integrated bioinformatics, a total of 43 biomarker genes were selected for further pathway enrichment analysis and correlation analysis, as well as diagnostic analysis. GO analysis showed these 43 TEPs mRNAs were mostly involved in protein binding, extracellular matrix and cellular protein metabolic process. KEGG pathway

enrichment analysis revealed that genes were mainly enriched in metabolic process. Eventually, by taking these 43 genes into spearman correlation analysis and receiving operating characteristic (ROC) curve analysis, we identified 1) TEPs RSL24D1 mRNA was negatively related to early pan cancer as compared to healthy controls and potential for early pan cancer diagnosis with a sensitivity of 71.8%, and a specificity of 64.3%; 2) HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2, and IFITM3 showed an upward trend in the expression from early to more advanced pan cancer stages, and had a diagnostic value for pan cancer with a sensitivity of 60.9%, 59.1%, 56.5%, 57.8%, 54.3%, 55.2%, 55.2% ,60.9%, and a specificity of 94.5%, 90.9%, 87.3%, 89.1%, 72.7%, 85.5%, 89.1%, 94.5%, respectively; 3) ARL2, FCGR2A, and KLHDC8B were positively associated with advanced, metastatic pan cancer as compared to healthy controls and could be diagnostic indicators for advanced pan cancer with a sensitivity of 59.2%, 61.8%, 59.7%, and a specificity of 80%, 89.1%, 83.6%, respectively. Therefore, our findings suggest that the 12-gene TEPs liquid-biopsy biomarkers will not only facilitate early diagnosis of pan cancer, but also be beneficial for pan cancer staging.

Keywords: tumor educated platelets; diagnosis; platelet mRNA; bioinformatics analysis

Background

Cancer is one of the leading causes of global mortality. In 2020, 1806,590 new cancer cases were reported with a death toll of 606,520 [1]. Cancer early detection is critical to prevent cancer from reaching an incurable state, and may improve cancer survival [2, 3]. Although the rapid development of medical technology, the diagnosis of cancer is still limited by some conditions [4, 5]. Nowadays, the diagnosis of cancer is based on the combination of clinical

features, biochemical tests, radiology, endoscopy, and pathology analysis of tumor tissue. Tissue biopsy, an invasive and expensive examination, still represents the gold standard for tumor diagnosis. In order to overcome limitations of tissue acquisition, the use of blood-based biomarkers or liquid biopsies has been proposed [6, 7]. To date, multiple blood-based biomarkers are being evaluated as the subject of research, including circulating nucleic acids, circulating tumor cells (CTCs), metabolic products, extracellular vesicles and circulating miRNAs [8-11]. In addition, tumor-educated platelets (TEPs) have also been engaged in blood-based cancer diagnostics [12-14].

Platelets are anucleate cells in the blood which contain a number of organelles including mitochondria and dense granules. Besides, they are packed with multiple RNA and proteins, which are received from megakaryocytes [15, 16]. Platelets, as a second most-abundant cell type in peripheral blood, are usually known to be the central regulator of thrombosis and hemostasis, which are generated from the cytoplasm of megakaryocytes [17]. Recent studies have demonstrated that platelets play essential roles in tumor progression and metastasis [18]. Platelets can be activated by tumor cells directly or indirectly, leading to their behavior and RNA profiles alteration. Thus, the tumor-educated platelets (TEPs) formed, which in turn, promoted the metastasis of tumor cells. TEPs RNAs have been emerging as potential blood-based biomarkers for cancer diagnosis, prognosis and prediction [12, 13]. In addition, TEPs RNAs are advantageous in cancer diagnosis as biomarkers due to convenience and simplicity of platelet collection. Furthermore, recent studies have proved that TIMP1 and TGA2B mRNA in TEPs, as well as a three-platelet mRNA set: MAX, MTURN and HLA-B et.al

are diagnostic biomarkers for colorectal cancer and lung cancer respectively [19-22]. However, these reports have some limitations, including limited clinical samples, lacked cancer stage prediction and through them, only one single cancer species could be predicted. These factors all limit the application of TEPs biomarkers in clinical screening, diagnosis and staging of pan cancer.

Next-generation sequencing (NGS) technology provides a new perspective for the study of pathological mechanisms underlying various cancers and has been widely applied to diagnose, assess and treat neoplasma [14, 23, 24]. Bioinformatics are new disciplines combined with life science and computer science which can be used to collect, process, store, disseminate, analyze/reanalyze and interpret genetic data.

Here, we downloaded the next-generation sequencing datasets, GSE68086 (285 samples) and GSE89843 (636 samples), from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). We analyzed these datasets deeply and explored the mechanisms of changes in TEPs RNA profile utilizing integrated bioinformatics analysis, correlation analysis and diagnostic analysis. Eventually, we found that RSL24D1, HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2, IFITM3, ARL2, FCGR2A and KLHDC8B in TEPs, which mainly enriched in protein binding, extracellular matrix and metabolic process, were negatively or positively correlated with pan cancer as compared to platelets of healthy controls, and they are potential for pan cancer diagnosis and staging.

Material and Methods

Collection and inclusion criteria of studies

We searched the GEO database for the following keywords: “cancer” (study keyword), “platelets” (study keyword), “Homo sapiens” (organism), “expression profiling by array or sequencing” (study type). The inclusion criteria for studies were as follows: (1) samples should be in three platelets groups including those collected from the healthy samples, early cancer samples and metastatic cancer samples, (2) the sample count should be proper, and (3) sufficient information should be used to perform the analysis. Consequently, GSE68086 and GSE89843 were collected for analyzing.

Microarray data

Two gene expression profiles (GSE68086 and GSE89843) were downloaded from the GEO database. The array data GSE68086 includes 285 platelets samples covering 55 healthy samples, 39 early, localized cancer samples and 191 advanced, metastatic cancer samples were collected for analysis. Similarly, the sequencing data GSE89843 chosen includes 636 samples covering 234 healthy samples, 55 early, localized cancer samples and 345 metastatic cancer samples. Details were shown in Figure1A and Figure2A.

Data processing

A large number of high-throughput data is stored in the GEO database for free research and a variety of research methods or tools are derived unavoidably. R is a language and environment for statistical computing and graphics, which provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests et. al) and graphical techniques. EdgeR package was used to identify the differential expression of genes by linear modelling. Genes with FC (fold change) > 1 and adj P value (adjusted P-value) < 0.05 were considered to

be differential expressed in platelets collected from early/metastatic cancer samples and healthy samples. Then, R software was quoted to obtain the heatmaps and volcano plots about differential expression of genes in TEPs.

Functional and pathway enrichment analysis

GO (Gene ontology analysis) is used to reveal the function of genes and gene products in any organism [25]. KEGG (Kyoto Encyclopedia of Genes and Genomes) allots genes and genomes functional meanings at the molecular and higher levels [26]. DAVID (Database for Annotation Visualization and Integrated Discovery, http://www.david.niaid.nih.gov) is a database which is applied to annotation, visualization, and integrated discovery [27]. Here we took advantage of DAVID to annotate and visualize. The P value < 0.1 was regarded as significantly enriched. Next, R software was quoted to obtain the graphics of the enriched pathways about differential expression of genes in TEPs.

PPI network construction and analysis of modules

The STRING (Search Tool for the Retrieval of Interacting Genes, https://string-db.org/) database aims to collect and integrate the information which represents all functional interactions between the expressed proteins through strengthening known and predicted protein-protein association data among plenty of organisms [28]. Thus protein–protein interaction (PPI) network of DEGs was built by STRING database. The Molecular Complex Detection (MCODE) based on Cytoscape was applied to screen modules of the PPI network with degree cutoff = 2, node score cutoff = 0.2, k-core = 2.

Statistical analysis

SPSS 23.0 software (SPSS, Chicago, IL, USA) was used to analyze statistics. Spearman

correlation coefficient was used to assess correlation between the average gene expression and that of the sample group for identifying genes whose expression change go up or down strictly monotonically with respect to group. The Mann Whitney test is then applied to identify the differential expressed genes among the different stages with $P < 0.05$ as the cutoff for the significance level. Receiving operating characteristic (ROC) curve analysis was used to evaluate the discriminatory power of the combinations. Data were shown as median ± inter quartile range. $P < 0.05$ were considered statistically significant and all tests were two-sided.

Results

Analysis of the mRNA profiles of TEPs from localized or metastatic pan cancer patients as compared to platelets from healthy controls

The next-generation sequencing data of GSE68086 has 285 platelet samples collected from healthy controls (n = 55), early, localized (n = 39) or advanced, metastatic cancer (n = 191) (Figure 1A). The early, localized cancer patient cohort includes five tumor types containing breast cancer (BrCa, n = 14), colorectal cancer (CRC, n = 6), hepatobiliary cancer (HBC, n = 4), non-small cell lung carcinoma (NSCLC, n = 4), and pancreatic cancer (PAAD, n = 11). The advanced, metastatic cancer patient cohort consists of six tumor types, including BrCa (n = 25), CRC (n = 36), glioblastoma (GBM, n = 40), HBC (n = 10), NSCLC (n = 56), and PAAD (n = 24) (Figure 1A). There were 3905 differential expressed mRNAs, among which 1060 were up-regulated and 2845 were down-regulated in early, localized cancer treated platelets as compared to platelet samples of healthy controls, whereas 3059 differential expressed mRNAs, among which 854 were up-regulated and 2205 were down-regulated in advanced, metastatic cancer treated platelets as compared to platelet samples of healthy controls (Figure 1 B-C,

supplementary figure 1 A-B, Table S1 and S2). To further investigate the differential expression between early, localized cancer treated platelets and advanced, metastatic cancer treated platelets, we explored commonly altered genes by using Venny 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/index.html). We verified 2593 commonly altered differential mRNAs from above platelets RNA-sequencing datasets, among them 669 were consistently up-regulated and 1924 were consistently down-regulated (Figure 1 D).

mRNA profiles of TEPs are distinct between localized and metastatic NSCLC cancer

The data of GSE89843 totally has 779 platelet samples collected from healthy controls (n = 234), individuals without reported cancers, but with inflammatory conditions (n = 143), early, localized (n = 57) or metastatic NSCLC cancer (n = 345), respectively. Here, we chose 636 of them excluding individuals with inflammatory conditions for further analysis (Figure 2A). Screening the differential expressed genes, a total of 164 differential expressed mRNAs were identified between 57 early, localized tumor educated platelets samples and 234 platelets samples of healthy controls, including 117 up-regulated and 47 down-regulated genes (Table S3). In addition, there were 49 differential expressed mRNAs between 402 metastatic NSCLC treated platelets samples and 234 platelets samples of healthy controls, including 39 up-regulated and 10 down-regulated genes (Table S4). Hierarchical clustering and a volcano plot were implemented to identify the differential expressed mRNAs (Figure 2 B-C, supplementary figure 2 A-B). For more insights into the differential expression between early, localized cancer treated platelets and metastatic cancer treated platelets, a venn diagram exhibited the common altered genes in Figure 2D. Among the 33 commonly altered differential mRNAs from above platelets RNA-sequencing datasets, 26 were consistently up-regulated

and 7 were consistently down-regulated.

Identification of differential expressed mRNAs in platelets between localized and metastatic pan cancer patients

To investigate the differential expressed mRNAs in platelets between localized or metastatic cancer and healthy donors, we analyzed the microarray data of GSE68068 and GSE89843. Using fold change (FC) ≥ 1 as the cut-off criterion, we extracted 3905 (1060 up-regulated and 2845 down-regulated) vs 164 (117 up-regulated and 47 down-regulated) DEGs (differential expressed genes) from localized tumor educated platelets as compared to platelet samples of healthy controls, and 3059 (854 up-regulated and 2205 down-regulated) vs 49 (39 up-regulated and 10 down-regulated) DEGs from metastatic tumor educated platelets in the two datasets respectively (Figure 1B-C, Figure 2B-C, TableS1-4). To further investigate the key differential genes between localized tumor educated platelets and metastatic tumor educated platelets, we integrated four groups of DEGs mentioned above to further take the intersection. Using available venn website, 20 common DEGs were identified, 74 DEGs were only in the localized tumor educated platelets, whereas 13 DEGs were only in the metastatic tumor educated platelets (supplementary figure 3, Table S5). In addition, to further screen out the consistent altered DEGs in the localized or/and metastatic tumor educated platelets, we integrated the four up-regulated/down-regulated groups to take the intersection. And we extracted 13 common DEGs, 8 DEGs which were only in the localized tumor educated platelets and 12 DEGs which were only in the metastatic tumor educated platelets in the four up-regulated groups, whereas 2 common DEGs, 7 DEGs which were only in the localized tumor educated platelets and 1 DEGs which was only in the metastatic tumor educated

platelets in the four down-regulated groups (Figure 3A-B, Table 1). The DEGs which were only in the localized tumor educated platelets (PLXNB3, SAMD14, ALAS2, C4orf48, PARP10, EHBP1L1, DYSF, SSBP4, LRRC75A, CD69, RSL24D1, ZNF667, PPT1, IARS, HERC3), only in the metastatic tumor educated platelets (FCGR2A, KLHDC8B, DEFA3, IGFBP2, MAOB, ZNF346, ARL2, MMP1, KLHL35, CA1, RP11-525A16.4, CTD-2509G16.2, MS4A1) and which were common in both localized and metastatic tumor educated platelets (HPSE, IFI27, LGALS3BP, CRYM, WASF1, HBD, COL6A3, PRSS50 , LAMB2, LTF, TPM2, TYMP, NELL2, SLC38A1, IFITM3) were taken for the further analysis.

Molecular concepts significantly enriched in tumor educated platelets

To explore the underlying mechanism and signaling pathways of these DEGs enriched in, DAVID and KEGG were performed to acquire functional and pathway enrichment analysis. GO (gene ontology) analysis was employed to functionally annotate the differential expressed platelet RNAs. The most significant output of GO analysis was related to protein binding, extra-cellular matrix, cellular protein metabolic process, mitochondrial outer membrane and innate immune response in mucosa, et. al (Figure 3C-D). For biological processes, 43 DEGs were enriched in negative regulation of viral genome replication, positive regulation of osteoblast proliferation, cellular protein metabolic process, innate immune response in mucosa, extracellular matrix organization, negative regulation of GTPase activity, and antibacterial humoral response. In addition, molecular function illustrated that they were involved in protein binding, and serine-type endopeptidase activity. Furthermore, cell component showed that they were enriched in extracellular exosome, extracellular region, proteinaceous extracellular matrix, extracellular matrix, mitochondrial outer membrane, and extracellular space (Figure 3D,

supplementary figure 4, Table S6). By using KEGG analysis, we found these 43 DEGs were significantly enriched in metabolic process, mostly in glycine, serine and threonine metabolism (Figure 3E, Table S7). Besides, PPI network consisted of 41 nodes and 51 edges with a confidence score of ≥0.15. The correlated degree scores constructed with Cytoscape MCODE were 3.826, 3.684, 3.667, 2.8, 2.769, and 2.75, indicating that the platelet-derived DEGs were interacted with each other (Figure 3F).

Diagnostic value of RSL24D1 for early, localized pan cancer

According to the above results, we have identified that TEPs DEGs, including PLXNB3, SAMD14, ALAS2, C4orf48, PARP10, EHBP1L1, DYSF, SSBP4, LRRC75A, CD69, RSL24D1, ZNF667, PPT1, IARS, and HERC3 which were in the localized cancer with four groups intersection. To further investigate the signature of these TEPs DEGs, we primarily checked whether some genes had their expression-level changes correlated with the early, localized pan cancer. By using Spearman correlation coefficient, we demonstrated PLXNB3 ($r = 0.359$, $p < 0.001$), SAMD14 ($r = 0.261$, $p = 0.011$), LRRC75A ($r = -0.592$, $p < 0.001$), CD69 ($r = -0.51$, $p < 0.001$), RSL24D1 ($r = -0.418$, $p < 0.001$), ZNF667 ($r = -0.331$, $p = 0.001$), IARS ($r = -0.286$, $p = 0.005$), and HERC3 ($r = -0.369$, $p < 0.001$) were positively or negatively correlated with the early pan cancer based on the data from GSE68086 (Figure 4A), whereas PLXNB3 ($r = 0.134$, $p = 0.022$), SAMD14 ($r = 0.291$, $p < 0.001$), ALAS2 ($r = 0.294$, $p < 0.001$), C4orf48 ($r = 0.249$, $p < 0.001$), PARP10 ($r = 0.192$, $p = 0.001$), EHBP1L1 ($r = 0.303$, $p < 0.001$), SSBP4 ($r = 0.212$, $p < 0.001$), LRRC75A ($r = -0.291$, $p < 0.001$), CD69 ($r = -0.191$, $p = 0.001$), RSL24D1 ($r = -0.296$, $p < 0.001$), ZNF667 ($r = -0.242$, $p < 0.001$), PPT1 ($r = -0.175$, $p = 0.003$), IARS ($r = -0.25$, $p < 0.001$), and HERC3 ($r = -0.26$, $p < 0.001$) were positively or negatively correlated with the early

NSCLC cancer based on the data from GSE89843 (Supplementary figure 5A). In addition, to further explore the diagnostic value of these TEPs DEGs, ROC analysis was carried out. Diagnostic analysis results showed that PLXNB3 (AUC=0.71, P=0.001), LRRC75A-AS1 (AUC = 0.847, P < 0.001), CD69 (AUC = 0.799, P < 0.001), RSL24D1 (AUC = 0.745, P < 0.001), and HERC3 (AUC = 0.716, P < 0.001) had a diagnostic value based on the data from GSE68086 (Figure 4B), whereas SAMD14 (AUC = 0.712, P < 0.001), ALAS2 (AUC = 0.714, P < 0.001), EHBP1L1 (AUC = 0.72, P < 0.001), LRRC75A-AS1 (AUC = 0.712, P < 0.001) and RSL24D1 (AUC = 0.715, P < 0.001) had a diagnostic value based on the data from GSE89843 (Supplementary figure 5B). Taking the results from two datasets into intersection, we identified RSL24D1 was negatively correlated with the early, localized cancer as compared to healthy controls, meanwhile it had a diagnostic value for early, localized cancer with a sensitivity of 71.8%, and a specificity of 64.3%.

Association of the TEPs DEGs signature with pan cancer diagnosis and staging

HPSE, IFI27, LGALS3BP, CRYM, WASF1, HBD, COL6A3, PRSS50, LAMB2, LTF, TPM2, TYMP, NELL2, SLC38A1, and IFITM3 were differentially expressed in both localized and metastatic cancer educated platelets as compared to platelets of healthy controls. By using Spearman correlation coefficient, we detected that in addition to TYMP, the expression levels of the remaining 14 TEPs DEGs were positively or negatively related to pan cancer stage on basis of the data from GSE68086 (Figure 5A), showing an upward or downward trend in the expression of TEPs DEGs from early to more advanced stages, whereas a total of these 15 TEPs DEGs were positively or negatively correlated with the stage of NSCLC on basis of the data from GSE89843 (Supplementary figure 6A). Additionally, diagnostic analysis results

indicated that HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2, and IFITM3 (AUC > 0.7, P < 0.001) had predictive validation value for pan cancer based on the data from GSE68086 and GSE89843 (Figure 5B, Supplementary figure 6B). Taking both correlation and diagnostic value into consideration, we verified HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2, and IFITM3 were positively related to pan cancer stage and were potential for pan cancer diagnosis and staging with a sensitivity of 60.9%, 59.1%, 56.5%, 57.8%, 54.3%, 55.2%, 55.2% ,60.9%, and a specificity of 94.5%, 90.9%, 87.3%, 89.1%, 72.7%, 85.5%, 89.1%, 94.5%, respectively.

Diagnostic value of ARL2, FCGR2A, and KLHDC8B for advanced, metastatic pan cancer

We have known that FCGR2A, KLHDC8B, DEFA3, IGFBP2, MAOB, ZNF346, ARL2, MMP1, KLHL35, CA1, RP11-525A16.4, CTD-2509G16.2, and MS4A1 were only differentially expressed in advanced, metastatic pan cancer. Here, we aimed to explore the diagnostic value of these TEPs DEGs for advanced, metastatic pan cancer. By using Spearman correlation coefficient, we identified that apart from CA1, the expression levels of the remaining 12 TEPs DEGs were positively or negatively correlated with advanced pan cancer through integrating two datasets from GSE68086 and GSE89843 (Figure 6A, Supplementary figure 7A). Besides, diagnostic analysis results showed that FCGR2A (AUC = 0.831, P < 0.001), KLHDC8B(AUC = 0.774, P < 0.001), MAOB (AUC = 0.77, P < 0.001), ZNF346 (AUC = 0.742, P < 0.001), ARL2 (AUC = 0.769, P < 0.001), MMP1 (AUC = 0.714, P < 0.001), and MS4A1 (AUC = 0.87, P < 0.001) had the diagnostic value based on the data from GSE68086 (Figure 6B). However, by analyzing GSE89843 data, we found only FCGR2A (AUC = 0.705, P < 0.001), KLHDC8B (AUC = 0.707, P < 0.001), IGFBP2 (AUC = 0.711, P < 0.001), and ARL2

(AUC = 0.791, P < 0.001) had the diagnostic value for metastatic NSCLC cancer (Supplementary figure 7B). Taking these results into intersection, we detected ARL2, FCGR2A, and KLHDC8B were negatively correlated with the advanced, metastatic pan cancer in comparison with healthy controls and they were essential for advanced, metastatic pan cancer diagnosis with a sensitivity of 59.2%, 61.8%, 59.7%, and a specificity of 80%, 89.1%, 83.6%, respectively.

Discussion

Diagnosis of cancer is difficult and complicated which mainly depends on clinical experience, patients' clinical manifestations and signs, laboratory examinations, imaging examinations, and histopathological examination at present. Imaging examinations including radiography, ultrasonography, computed tomography (CT) and magnetic resonance imaging (MRI) are usually used for cancer screening and monitoring. Tumor tissue biopsy is deemed to be the gold standard of cancer sub-typing. But it is expensive, invasive and even rather difficult to obtain in several cases [29]. Most of these traditional methods used to screen for cancer detect the disease already at a later stage [30]. Mammography is the standard imaging modality for early detection of breast cancer. However, women with dense breast tissue, mammography is less sensitive and not all breast cancers can be detected [31]. Colonoscopy and bronchoscope are effective methods for diagnosing colorectal and lung cancers respectively [32, 33]. However, they usually detect colorectal or lung cancer at an advanced stage. Hepatobiliary cancer and glioblastoma are with high malignancy, morbidity and mortality rates, and low cure rates. If they are diagnosed at an early stage, these current situations will be improved. Recently, many efforts are being made to improve the early diagnostic methods,

320 noninvasive liquid biopsies, especially popular in the diagnosis of solid tumors which are
321 difficult to obtain tissue biopsy before surgery, offering a promising alternative for tumor
322 diagnosis, prediction and monitoring [34]. Comfortingly, compared to traditional tissue biopsy,
323 liquid biopsies are more advantages, since: 1) low side effects, non-invasive, and being
324 capable of repeatable sampling, 2) not relying on imaging examinations, 3) effectively dealing
325 with tumor heterogeneity. In this study, we identified 12-gene TEPs mRNAs as liquid-biopsy
326 biomarkers for diagnosing and staging pan cancer including breast cancer, colorectal cancer,
327 lung cancer, hepatobiliary cancer, glioblastoma and pancreatic cancer.

328 Blood samples for liquid biopsy include circulating tumor cells (CTCs), cell-free nucleic acids,
329 exosomes (DNA, RNA, miRNA, proteins), and TEPs [35]. Liquid biopsy shows bright future
330 prospects for early detection of cancers, but CTCs, cell-free nucleic acids and exosomes have
331 limitations of being very low in number in early stages of multiple cancers [35, 36].
332 Platelets-associated indicators and platelet counts have been shown associated with the
333 prognosis of resectable lung and colorectal cancers [37-39]. In addition to the predictive role of
334 platelets-associated indicators in cancer, platelet counts, or protein markers, Best et. al [13]
335 have identified RNA profiles of TEPs can be used to diagnose cancer patients with 96%
336 accuracy through NGS based on 283 blood platelet samples, isolated from healthy controls
337 and patients with cancer. Furthermore, recent studies have demonstrated that TIMP1 and
338 TGA2B mRNA in tumor-educated platelets are diagnostic biomarkers for colorectal cancer and
339 lung cancer respectively [19, 20]. In this study, integrating the NGS datasets, GSE68086 and
340 GSE89843, downloaded from GEO datasets, we identified 43 TEPs genes were consistently

differential expressed through bioinformatic analysis, and they were involved in many biological processes including protein binding, extracellular matrix, cellular protein metabolic process, mitochondrial outer membrane and innate immune response in mucosa, and enriched in multiple pathways, such as glycine, serine and threonine metabolism. After further investigating these 43 DEGs through correlation and diagnostic analysis, we finally demonstrated that 12 DEGs, including RSL24D1, HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2, IFITM3, ARL2, FCGR2A, and KLHDC8B in TEPs were essential for pan cancer detection at different stages.

RSL24D (ribosomal L24 domain containing 1), also called C15orf15, RPL24L, or My024, encodes probable ribosome biogenesis protein RLP24, involved in the biogenesis of the 60S ribosomal subunit, ensuring the docking of GTPBP4/NOG1 to pre-60S particles. Only few studies have been reported on RSL24D1 at present. It is related to hypercholesterolemia, children chronic kidney disease (CKD) when compared with control samples [40, 41]. A recent study has shown that change of RSL24D1 is associated with advanced-stage NSCLC compared with stage I NSCLC through genome-wide methylation profiles analysis, which is the only one mentioned that RSL24D1 is related to tumor [42]. In our study, we found RSL24D1 in TEPs was negatively associated with early pan cancer, including breast cancer, lung cancer, CRC, PAAD, HBC, as compared to healthy controls. In addition, we also demonstrated the diagnostic value of RSL24D1 for early pan cancer with a sensitivity of 71.8%, and a specificity of 64.3%. Furthermore, we demonstrated that ARL2, FCGR2A, and KLHDC8B in TEPs were positively related to metastatic pan cancer in comparison with healthy

controls and had potential diagnostic significance for metastatic pan cancer with a sensitivity of 59.2%, 61.8%, 59.7%, and a specificity of 80%, 89.1%, 83.6%, respectively. Although the role of KLHDC8B in tumors is not clear yet, excessive studies have reported that ARL2 and FCGR2A are associated with cancer growth, invasion, as well as recurrence [43, 44].

In addition to the value of RSL24D1 in TEPs for early pan caner, as well as ARL2, FCGR2A, and KLHDC8B for metastatic pan cancer, we also identified that HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2 and IFITM3 in TEPs were positively related to pan cancer stage and essential for diagnosing pan cancer with a sensitivity of 60.9%, 59.1%, 56.5%, 57.8%, 54.3%, 55.2%, 55.2% ,60.9%, and a specificity of 94.5%, 90.9%, 87.3%, 89.1%, 72.7%, 85.5%, 89.1%, 94.5%, respectively. Ketimine reductase mu-crystallin and hemoglobin subunit delta which are encoded by CRYM and HBD respectively, have not yet been found to play a role in tumor progression. However multiple studies have shown the remaining genes are closely related to tumor progression. HPSE over-expression is related to tumor growth, metastasis, and angiogenesis [45]. HPSE knockdown suppresses the tumor growth and lung metastasis of melanoma [46]. IFI27 (interferon alpha-inducible protein 27) has been involved in different apoptosis signaling pathways, including type-I interferon-induced apoptosis and TNFSF10-induced apoptosis [47-49], and the innate immune response [50, 51]. In addition, up-regulation of IFI27 participates in the invasion and proliferation of many types of cancers, such as oral squamous cell carcinoma, pancreatic ductal adenocarcinoma and breast cancer [52-54]. LGALS3BP encodes galectin-3-binding protein which is involved in regulating cell adhesion, growth, differentiation, apoptosis, as well as angiogenesis. Its up-regulation has a

prognostic value for multiple cancers, such as breast cancer, gastric cancer, colorectal cancer, liver cancer and GBM [55-59]. COL6A3 has been demonstrated to have potential clinical significance of diagnosis or promotion roles in a number of cancers, including PAAD, CRC, and gynecologic oncology [60-63]. LAMB2, subunit beta-2 of laminin, mediates the attachment, migration and organization of various cells, including cancer cells, into tissues during embryonic development by interacting with other extracellular matrix components [64]. In addition, IFITM3P plays roles in promotion of both blood and solid tumors through different signaling pathways [65-67]. According to an online website, GEPIA (Gene Expression Profiling Interactive Analysis, http://gepia.cancer-pku.cn/), although RSL24D1, IFI27, LGALS3BP, LAMB2 and KLHDC8B are not correlated with prognosis of pan cancer patients, the expression of HPSE, CRYM, HBD, IFITM3, ARL2 and FCGR2A in tumor issues is negatively related to overall survival ($p<0.001$) in pan cancer (BrCa, CRC, HBC, NSCLC, PAAD, GBM) patients (data not shown), which is consistent with our research in TEPs. However, by using GEPIA, we found the higher expression of COL6A3 leads to the longer overall survival ($p<0.001$), which was opposite to previous studies, and our findings demonstrated that COL6A3 showed an upward trend in the expression from early to more advanced pan cancer stages. Different clinical sample sources based on the website may explain this phenomenon.

Self-sufficiency in growth signals, insensitivity to antigrowth signals, limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis, avoiding immune destruction, tumor promotion inflammation, deregulating cellular energetics, genome instability and mutation and resisting cell death are hallmarks of cancer [68]. According to the GO term

404 analysis and KEGG pathway analysis, we found these TEPs mRNAs were correlated with
405 protein binding, extracellular matrix, cellular protein metabolic process, mitochondrial outer
406 membrane and innate immune response in mucosa, as well as enriched in metabolic process,
407 mostly glycine, serine and threonine metabolism. Metabolic abnormalities, as one of the ten
408 hallmarks of cancer, are mutually causal with tumor tumorigenesis. Tumor blood metastasis
409 can be divided into three steps including the translocation of vascular endothelial cells from
410 tumor cells through the tumor tissue from the primary site, tumor cells rolling with the blood, as
411 well as the implantation of tumor cells in the metastatic site. And multiple cell adhesion
412 molecules, extracellular matrix and other blood cells are involved in tumor metastatic process.
413 Recently, platelets have been found to have a critical role in the promotion of tumorigenesis
414 during the metastasization. Mechanisms may be as follows: 1) Platelets can be activated by
415 tumor cells in the blood vessels, then aggregating around tumor cells to form a tumor
416 thrombus, thereby protecting the tumor cells from the immune system attack; 2) Platelets can
417 adhere to endothelial cells and tumor cells meanwhile through surface adhesion molecules,
418 such as p-selectin, which can act as a bridge between tumor cells and endothelial cells,
419 thereby helping tumor cells to adhere to the vascular endothelial cells at the metastatic site; 3)
420 Platelets can secrete a variety of biological factors, then promote tumor growth and
421 angiogenesis of tumor tissue [69, 70]. Taking these into consideration, we hypothesize that
422 cancer cells alter the expression of platelet mRNA, which leads to change of platelet function
423 and metabolic dysfunction, which in turn further affecting tumor progression. Thus, we believe
424 alternative TEPs mRNAs, including RSL24D1, HPSE, IFI27, LGALS3BP, CRYM, HBD,
425 COL6A3, LAMB2, IFITM3, ARL2, FCGR2A, and KLHDC8B mRNA can potentially serve as

non-invasive biomarkers for diagnosing pan cancer, even can predict the prognosis of pan cancer, although more scientific researches and evidences are needed to verify.

In conclusion, the present study firstly demonstrates that RSL24D1, HPSE, IFI27, LGALS3BP, CRYM, HBD, COL6A3, LAMB2, IFITM3, ARL2, FCGR2A, and KLHDC8B mRNA in TEPs can not only serve as non-invasive biomarkers for diagnosing pan cancer, but also can predict stages of pan cancer.

List of abbreviations

TEPs, tumor educated platelets;

DEGs, differentially expressed genes;

GO, gene ontology;

KEGG, kyoto encyclopedia of genes and genomes;

ROC, receiving operating characteristic;

CTCs, circulating tumor cells;

NGS, next-generation sequencing;

GEO, gene expression omnibus;

DAVID, database for annotation visualization and integrated discovery;

STRING, search tool for the retrieval of interacting genes;

PPI, protein–protein interaction;

MCODE, molecular complex detection;

BrCa, breast cancer;

CRC, colorectal cancer;

HBC, hepatobiliary cancer;

NSCLC, non-small cell lung carcinoma;

PAAD, pancreatic cancer;

GBM, glioblastoma;

CT, computed tomography;

MRI, magnetic resonance imaging.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets in this study are available from GEO database (https://www.ncbi.nlm.nih.gov/geo/) using accessions numbers GSE68086 and GSE89843

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by Key Program of the National Natural Science Foundation of China Grant 81130008 (to K.D.)

Authors' contributions

Xinxin Ge designed and performed the research, collected and analyzed data, and wrote the paper; Liuxia Yuan, Ying Hu and Bin Cheng helped to analyze data; Khan Muhammad Shoaib

helped to review the paper; Kesheng Dai initiated and supervised the project, designed research, analyzed and interpreted results, and revised the paper. All authors read and approved the final manuscript.

References

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. CA Cancer J Clin 2020, 70(1):7-30.
2. Jodal HC, Helsingen LM, Anderson JC, Lytvyn L, Vandvik PO, Emilsson L: Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. BMJ Open 2019, 9(10):e032773.
3. Yu G: Near-infrared diffuse correlation spectroscopy in cancer diagnosis and therapy monitoring. J Biomed Opt 2012, 17(1):010901.
4. Wang S, Xing C, Wu H, Dai M, Zhao Y: Pancreatic schwannoma mimicking pancreatic cystadenoma: A case report and literature review of the imaging features. Medicine (Baltimore) 2019, 98(24):e16095.
5. Asano N, Matsuzaki J, Ichikawa M, Kawauchi J, Takizawa S, Aoki Y, Sakamoto H, Yoshida A, Kobayashi E, Tanzawa Y et al: A serum microRNA classifier for the diagnosis of sarcomas of various histological subtypes. Nat Commun 2019, 10(1):1299.
6. Nalejska E, Maczynska E, Lewandowska MA: Prognostic and predictive biomarkers: tools in personalized oncology. Mol Diagn Ther 2014, 18(3):273-284.
7. Haber DA, Velculescu VE: Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer Discov 2014, 4(6):650-661.

8. Pos O, Biro O, Szemes T, Nagy B: Circulating cell-free nucleic acids: characteristics and applications. Eur J Hum Genet 2018, 26(7):937-945.

9. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C: Molecular analysis of circulating tumour cells-biology and biomarkers. Nat Rev Clin Oncol 2014, 11(3):129-144.

10. Qi Y, Cui L, Ge Y, Shi Z, Zhao K, Guo X, Yang D, Yu H, Cui L, Shan Y et al: Altered serum microRNAs as biomarkers for the early diagnosis of pulmonary tuberculosis infection. BMC Infect Dis 2012, 12:384.

11. Adhami M, Haghdoost AA, Sadeghi B, Malekpour Afshar R: Candidate miRNAs in human breast cancer biomarkers: a systematic review. Breast Cancer 2018, 25(2):198-205.

12. Best MG, Vancura A, Wurdinger T: Platelet RNA as a circulating biomarker trove for cancer diagnostics. J Thromb Haemost 2017, 15(7):1295-1306.

13. Best MG, Sol N, Kooi I, Tannous J, Westerman BA, Rustenburg F, Schellen P, Verschueren H, Post E, Koster J et al: RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. Cancer Cell 2015, 28(5):666-676.

14. Best MG, Sol N, In 't Veld S, Vancura A, Muller M, Niemeijer AN, Fejes AV, Tjon Kon Fat LA, Huis In 't Veld AE, Leurs C et al: Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets. Cancer Cell 2017, 32(2):238-252 e239.

15. Wurdinger T, In 't Veld S, Best MG: Platelet RNA as Pan-Tumor Biomarker for Cancer

513 Detection. Cancer Res 2020, 80(7):1371-1373.
514 16. Nilsson RJ, Balaj L, Hulleman E, van Rijn S, Pegtel DM, Walraven M, Widmark A,
515 Gerritsen WR, Verheul HM, Vandertop WP et al: Blood platelets contain tumor-derived
516 RNA biomarkers. Blood 2011, 118(13):3680-3683.
517 17. George JN: Platelets. Lancet 2000, 355(9214):1531-1539.
518 18. Nash GF, Turner LF, Scully MF, Kakkar AK: Platelets and cancer. Lancet Oncol 2002,
519 3(7):425-430.
520 19. Yang L, Jiang Q, Li DZ, Zhou X, Yu DS, Zhong J: TIMP1 mRNA in tumor-educated
521 platelets is diagnostic biomarker for colorectal cancer. Aging (Albany NY) 2019,
522 11(20):8998-9012.
523 20. Xing S, Zeng T, Xue N, He Y, Lai YZ, Li HL, Huang Q, Chen SL, Liu WL: Development
524 and Validation of Tumor-educated Blood Platelets Integrin Alpha 2b (ITGA2B) RNA for
525 Diagnosis and Prognosis of Non-small-cell Lung Cancer through RNA-seq. Int J Biol
526 Sci 2019, 15(9):1977-1992.
527 21. Liu L, Song X, Li X, Xue L, Ding S, Niu L, Xie L, Song X: A three-platelet mRNA set:
528 MAX, MTURN and HLA-B as biomarker for lung cancer. J Cancer Res Clin Oncol
529 2019, 145(11):2713-2723.
530 22. Sheng M, Dong Z, Xie Y: Identification of tumor-educated platelet biomarkers of
531 non-small-cell lung cancer. Onco Targets Ther 2018, 11:8143-8151.
532 23. Best MG, In 't Veld S, Sol N, Wurdinger T: RNA sequencing and swarm
533 intelligence-enhanced classification algorithm development for blood-based disease
534 diagnostics using spliced blood platelet RNA. Nat Protoc 2019, 14(4):1206-1234.

24. Stratton MR, Campbell PJ, Futreal PA: The cancer genome. Nature 2009, 458(7239):719-724.

25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000, 25(1):25-29.

26. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017, 45(D1):D353-D361.

27. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003, 4(5):P3.

28. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P et al: The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017, 45(D1):D362-D368.

29. Vaidyanathan R, Soon RH, Zhang P, Jiang K, Lim CT: Cancer diagnosis: from tumor to liquid biopsy and beyond. Lab Chip 2018, 19(1):11-34.

30. Kumar S, Srinivasan A, Nikolajeff F: Role of Infrared Spectroscopy and Imaging in Cancer Diagnosis. Curr Med Chem 2018, 25(9):1055-1072.

31. Niell BL, Freer PE, Weinfurtner RJ, Arleo EK, Drukteinis JS: Screening for Breast Cancer. Radiol Clin North Am 2017, 55(6):1145-1162.

32. Kim SY, Kim HS, Park HJ: Adverse events related to colonoscopy: Global trends and

future challenges. World J Gastroenterol 2019, 25(2):190-204.

33. Sumi T, Ikeda T, Sawai T, Shijubou N, Kure K, Yamada Y, Nakata H, Mori Y, Takahashi H: Comparison of ultrathin bronchoscopy with conventional bronchoscopy for the diagnosis of peripheral lung lesions without virtual bronchial navigation. Respir Investig 2020.

34. San Lucas FA, Allenson K, Bernard V, Castillo J, Kim DU, Ellis K, Ehli EA, Davies GE, Petersen JL, Li D et al: Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes. Ann Oncol 2016, 27(4):635-641.

35. Speicher MR, Pantel K: Tumor signatures in the blood. Nat Biotechnol 2014, 32(5):441-443.

36. Asghar S, Waqar W, Umar M, Manzoor S: Tumor educated platelets, a promising source for early detection of hepatocellular carcinoma: Liquid biopsy an alternative approach to tissue biopsy. Clin Res Hepatol Gastroenterol 2020.

37. Qian W, Ge XX, Wu J, Gong FR, Wu MY, Xu MD, Lian L, Wang WJ, Li W, Tao M: Prognostic evaluation of resectable colorectal cancer using platelet-associated indicators. Oncol Lett 2019, 18(1):571-580.

38. Wang JJ, Wang YL, Ge XX, Xu MD, Chen K, Wu MY, Gong FR, Tao M, Wang WJ, Shou LM et al: Prognostic Values of Platelet-Associated Indicators in Resectable Lung Cancers. Technol Cancer Res Treat 2019, 18:1533033819837261.

39. Sylman JL, Boyce HB, Mitrugno A, Tormoen GW, Thomas IC, Wagner TH, Lee JS, Leppert JT, McCarty OJT, Mallick P: A Temporal Examination of Platelet Counts as a

579 Predictor of Prognosis in Lung, Prostate, and Colon Cancer Patients. Sci Rep 2018,
580 8(1):6564.
581 40. Wuttke M, Wong CS, Wuhl E, Epting D, Luo L, Hoppmann A, Doyon A, Li Y,
582 Consortium CK, Sozeri B et al: Genetic loci associated with renal function measures
583 and chronic kidney disease in children: the Pediatric Investigation for Genetic Factors
584 Linked with Renal Progression Consortium. Nephrol Dial Transplant 2016,
585 31(2):262-269.
586 41. Li G, Wu XJ, Kong XQ, Wang L, Jin X: Cytochrome c oxidase subunit VIIb as a
587 potential target in familial hypercholesterolemia by bioinformatical analysis. Eur Rev
588 Med Pharmacol Sci 2015, 19(21):4139-4145.
589 42. Hong Y, Choi HM, Cheong HS, Shin HD, Choi CM, Kim WJ: Epigenome-Wide
590 Association Analysis of Differentially Methylated Signals in Blood Samples of Patients
591 with Non-Small-Cell Lung Cancer. J Clin Med 2019, 8(9).
592 43. Beghin A, Belin S, Hage-Sleiman R, Brunet Manquat S, Goddard S, Tabone E,
593 Jordheim LP, Treilleux I, Poupon MF, Diaz JJ et al: ADP ribosylation factor like 2 (Arl2)
594 regulates breast tumor aggressivity in immunodeficient mice. PLoS One 2009,
595 4(10):e7478.
596 44. Magnes T, Melchardt T, Hufnagl C, Weiss L, Mittermair C, Neureiter D, Klieser E,
597 Rinnerthaler G, Roesch S, Gaggl A et al: The influence of FCGR2A and FCGR3A
598 polymorphisms on the survival of patients with recurrent or metastatic squamous cell
599 head and neck cancer treated with cetuximab. Pharmacogenomics J 2018,
600 18(3):474-479.

601 45. Masola V, Zaza G, Gambaro G, Franchi M, Onisto M: Role of heparanase in tumor
602 progression: Molecular aspects and therapeutic options. Semin Cancer Biol 2020,
603 62:86-98.
604 46. Liu XY, Tang QS, Chen HC, Jiang XL, Fang H: Lentiviral miR30-based RNA
605 interference against heparanase suppresses melanoma metastasis with lower liver
606 and lung toxicity. Int J Biol Sci 2013, 9(6):564-577.
607 47. Rosebeck S, Leaman DW: Mitochondrial localization and pro-apoptotic effects of the
608 interferon-inducible protein ISG12a. Apoptosis 2008, 13(4):562-572.
609 48. Gytz H, Hansen MF, Skovbjerg S, Kristensen AC, Horlyck S, Jensen MB, Fredborg M,
610 Markert LD, McMillan NA, Christensen EI et al: Apoptotic properties of the type 1
611 interferon induced family of human mitochondrial membrane ISG12 proteins. Biol Cell
612 2017, 109(2):94-112.
613 49. Liu N, Zuo C, Wang X, Chen T, Yang D, Wang J, Zhu H: miR-942 decreases
614 TRAIL-induced apoptosis through ISG12a downregulation and is regulated by AKT.
615 Oncotarget 2014, 5(13):4959-4971.
616 50. Xue B, Yang D, Wang J, Xu Y, Wang X, Qin Y, Tian R, Chen S, Xie Q, Liu N et al:
617 ISG12a Restricts Hepatitis C Virus Infection through the Ubiquitination-Dependent
618 Degradation Pathway. J Virol 2016, 90(15):6832-6845.
619 51. Chen Y, Jiao B, Yao M, Shi X, Zheng Z, Li S, Chen L: ISG12a inhibits HCV replication
620 and potentiates the anti-HCV activity of IFN-alpha through activation of the Jak/STAT
621 signaling pathway independent of autophagy and apoptosis. Virus Res 2017,
622 227:231-239.

52. Wang H, Qiu X, Lin S, Chen X, Wang T, Liao T: Knockdown of IFI27 inhibits cell proliferation and invasion in oral squamous cell carcinoma. World J Surg Oncol 2018, 16(1):64.

53. Chang X, Yang MF, Fan W, Wang LS, Yao J, Li ZS, Li DF: Bioinformatic Analysis Suggests That Three Hub Genes May Be a Vital Prognostic Biomarker in Pancreatic Ductal Adenocarcinoma. J Comput Biol 2020.

54. Szekely B, Bossuyt V, Li X, Wali VB, Patwardhan GA, Frederick C, Silber A, Park T, Harigopal M, Pelekanou V et al: Immunological differences between primary and metastatic breast cancer. Ann Oncol 2018, 29(11):2232-2239.

55. Song L, Tang JW, Owusu L, Sun MZ, Wu J, Zhang J: Galectin-3 in cancer. Clin Chim Acta 2014, 431:185-191.

56. Ilmer M, Mazurek N, Gilcrease MZ, Byrd JC, Woodward WA, Buchholz TA, Acklin K, Ramirez K, Hafley M, Alt E et al: Low expression of galectin-3 is associated with poor survival in node-positive breast cancers and mesenchymal phenotype in breast cancer stem cells. Breast Cancer Res 2016, 18(1):97.

57. Cheng D, Liang B, Li Y: Serum galectin-3 as a potential marker for gastric cancer. Med Sci Monit 2015, 21:755-760.

58. Tas F, Bilgin E, Tastekin D, Erturk K, Duranyildiz D: Clinical Significance of Serum Galectin-3 Levels in Gastric Cancer Patients. J Gastrointest Cancer 2016, 47(2):182-186.

59. He X, Zhang S, Chen J, Li D: Increased LGALS3 expression independently predicts shorter overall survival in patients with the proneural subtype of glioblastoma. Cancer

Med 2019, 8(5):2031-2040.

60. Svoronos C, Tsoulfas G, Souvatzi M, Chatzitheoklitos E: Prognostic value of COL6A3 in pancreatic adenocarcinoma. Ann Hepatobiliary Pancreat Surg 2020, 24(1):52-56.

61. Kang CY, Wang J, Axell-House D, Soni P, Chu ML, Chipitsyna G, Sarosiek K, Sendecki J, Hyslop T, Al-Zoubi M et al: Clinical significance of serum COL6A3 in pancreatic ductal adenocarcinoma. J Gastrointest Surg 2014, 18(1):7-15.

62. Liu W, Li L, Ye H, Tao H, He H: Role of COL6A3 in colorectal cancer. Oncol Rep 2018, 39(6):2527-2536.

63. Sherman-Baust CA, Weeraratna AT, Rangel LB, Pizer ES, Cho KR, Schwartz DR, Shock T, Morin PJ: Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. Cancer Cell 2003, 3(4):377-386.

64. Qin Y, Rodin S, Simonson OE, Hollande F: Laminins and cancer stem cells: Partners in crime? Semin Cancer Biol 2017, 45:3-12.

65. Liu Y, Lu R, Cui W, Pang Y, Liu C, Cui L, Qian T, Quan L, Dai Y, Jiao Y et al: High IFITM3 expression predicts adverse prognosis in acute myeloid leukemia. Cancer Gene Ther 2020, 27(1-2):38-44.

66. Liu X, Chen L, Fan Y, Hong Y, Yang X, Li Y, Lu J, Lv J, Pan X, Qu F et al: IFITM3 promotes bone metastasis of prostate cancer cells by mediating activation of the TGF-beta signaling pathway. Cell Death Dis 2019, 10(7):517.

67. Min J, Feng Q, Liao W, Liang Y, Gong C, Li E, He W, Yuan R, Wu L: IFITM3 promotes hepatocellular carcinoma invasion and metastasis by regulating MMP9 through

p38/MAPK signaling. FEBS Open Bio 2018, 8(8):1299-1311.

68. Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. Cell 2011, 144(5):646-674.

69. Contursi A, Grande R, Dovizio M, Bruno A, Fullone R, Patrignani P: Platelets in cancer development and diagnosis. Biochem Soc Trans 2018, 46(6):1517-1527.

70. Schlesinger M: Role of platelets and platelet receptors in cancer metastasis. J Hematol Oncol 2018, 11(1):125.

674 **Table I** Consistent altered DEGs in the localized or / and metastatic tumor educated platelets

Types	Down	Up
early/localized cancer	LRRC75A-AS1	PLXNB3
	CD69	SAMD14
	RSL24D1	ALAS2
	ZNF667-AS1	C4orf48
	PPT1	PARP10
	IARS	EHBP1L1
	HERC3	DYSF
		SSBP4
common between early and advanced cancer	NELL2	HBD
	SLC38A1	IFITM3
		CRYM
		IFI27
		HPSE
		WASF1
		LGALS3BP
		TYMP
		COL6A3
		LAMB2
		TPM2
		LTF
		PRSS50
advanced cancer	MS4A1	FCGR2A
		KLHDC8B
		DEFA3
		IGFBP2
		MAOB
		ZNF346
		ARL2
		MMP1
		KLHL35
		CA1
		RP11-525A16.4
		CTD-2509G16.2

675 **Figure legends**

676 **Figure 1** Identification of DEGs in TEPs from localized or metastatic pan cancer patients in
677 comparation with platelets from healthy controls based on the datasets GSE68086. **(A)**,
678 Number of platelet samples of healthy controls and cancer patients with different stages or
679 types of cancer based on GSE68086. **(B-C)**, Hierarchical clustering heatmap of DEGs in the
680 expression profiling datasets GSE68086. **(B)**, Heatmap of DEGs in TEPs collected from
681 healthy controls and early, localized cancer patients. **(C)**, heatmap of DEGs in TEPs collected
682 from healthy controls and advanced, metastatic cancer patients. The horizontal axis indicates
683 the sample, and the vertical axis indicates the DEGs. Red represents the up-regulated DEGs
684 and green represents the down-regulated DEGs. DEGs, differential expressed genes. **(D)**,
685 Identification of TEPs mRNAs between localized and metastatic pan cancer. Left, commonly
686 altered differential expressed TEPs mRNAs. Middle, identification of up-regulated differential
687 expressed TEPs mRNAs. Right, identification of down-regulated differential expressed TEPs
688 mRNAs.

689 **Figure 2** mRNA profiles of TEPs from localized or metastatic NSCLC cancer patients as
690 compared to platelets from healthy controls based on the datasets GSE89843. **(A)**, Number of
691 platelet samples of healthy controls and NSCLC cancer patients at different stages. **(B-C)**,
692 Hierarchical clustering heatmap of DEGs in the expression profiling datasets GSE89843. **(B)**,
693 Heatmap of DEGs in TEPs collected from healthy controls and early, localized NSCLC cancer
694 patients. **(C)**, heatmap of DEGs in TEPs collected from healthy controls and metastatic
695 NSCLC cancer patients. The horizontal axis indicates the sample, and the vertical axis
696 indicates the DEGs. Red represents the up-regulated DEGs and green represents the

down-regulated DEGs. **(D)**, Identification of TEPs mRNAs between localized and metastatic NSCLC cancer. Left, commonly altered differential expressed TEPs mRNAs. Middle, identification of up-regulated differential expressed TEPs mRNAs. Right, identification of down-regulated differential expressed TEPs mRNAs.

Figure 3 Analysis of the spliced RNA repertoire of TEPs from pan cancer patients at different stages. **(A-B)**, Identification of DEGs in the four datasets (GSE68086 early pan cancer and metastatic pan cancer, GSE89843 early NSCLC cancer and metastatic NSCLC cancer) via Venn diagrams software. **(A)**, identification of up-regulated differential expressed TEPs mRNAs. **(B)**, identification of down-regulated differential expressed TEPs mRNAs. Different colors represent different datasets. **(C-D)**, GO analyses of the DEGs according to their biological process, cellular component and molecular function. GO, gene ontology. **(E)**, KEGG pathway enrichment analysis. Dot size represents the number of genes in each KEGG pathway; P-value: Red < purple < blue. KEGG, Kyoto Encyclopedia of Genes and Genomes. **(F)** Protein-protein interaction network of DEGs visualized through String datasets.

Figure 4 Diagnostic value of TEPs DEGs for early, localized pan cancer based on the datasets GSE68086. **(A)**, Correlation analysis between expression levels of 15 TEPS DEGs and two groups including healthy control groups and early cancer groups. HC, healthy control; EC, early cancer; The 15 TEPS DEGs are PLXNB3, SAMD14, ALAS2, C4orf48, PARP10, EHBP1L1, DYSF, SSBP4, LRRC75A, CD69, RSL24D1, ZNF667, PPT1, IARS, and HERC3, respectively. **(B)**, ROC analysis of sensitivity and specificity of the above 15 TEPS DEGs signature in predicting the diagnosis of early pan cancer patients.

Figure 5 Identification of 15 TEPS DEGs signatures for pan cancer diagnosis based on the

datasets GSE68086. **(A)**, Correlation analysis between expression levels of 15 TEPS DEGs and three groups including healthy control groups, early cancer groups and advanced cancer groups. HC, healthy control; EC, early cancer; AC, advanced cancer. The 15 TEPS DEGs are HPSE, IFI27, LGALS3BP, CRYM, WASF1, HBD, COL6A3, PRSS50, LAMB2, LTF, TPM2, TYMP, NELL2, SLC38A1, and IFITM3, respectively. **(B)**, ROC analysis of sensitivity and specificity of the above 15 TEPS DEGs signature in predicting the diagnosis of pan cancer patients.

Figure 6 Diagnostic value of 13 TEPs DEGs for advanced, metastatic pan cancer based on the datasets GSE68086. **(A)**, Correlation analysis between expression levels of 13 TEPS DEGs and two groups including healthy control groups and advanced cancer groups. HC, healthy control; AC, advanced cancer. The 13 TEPS DEGs are FCGR2A, KLHDC8B, DEFA3, IGFBP2, MAOB, ZNF346, ARL2, MMP1, KLHL35, CA1, RP11-525A16.4, CTD-2509G16.2, and MS4A1, respectively. **(B)**, ROC analysis of sensitivity and specificity of the above 15 TEPS DEGs signature in predicting the diagnosis of advanced pan cancer patients.

Figures

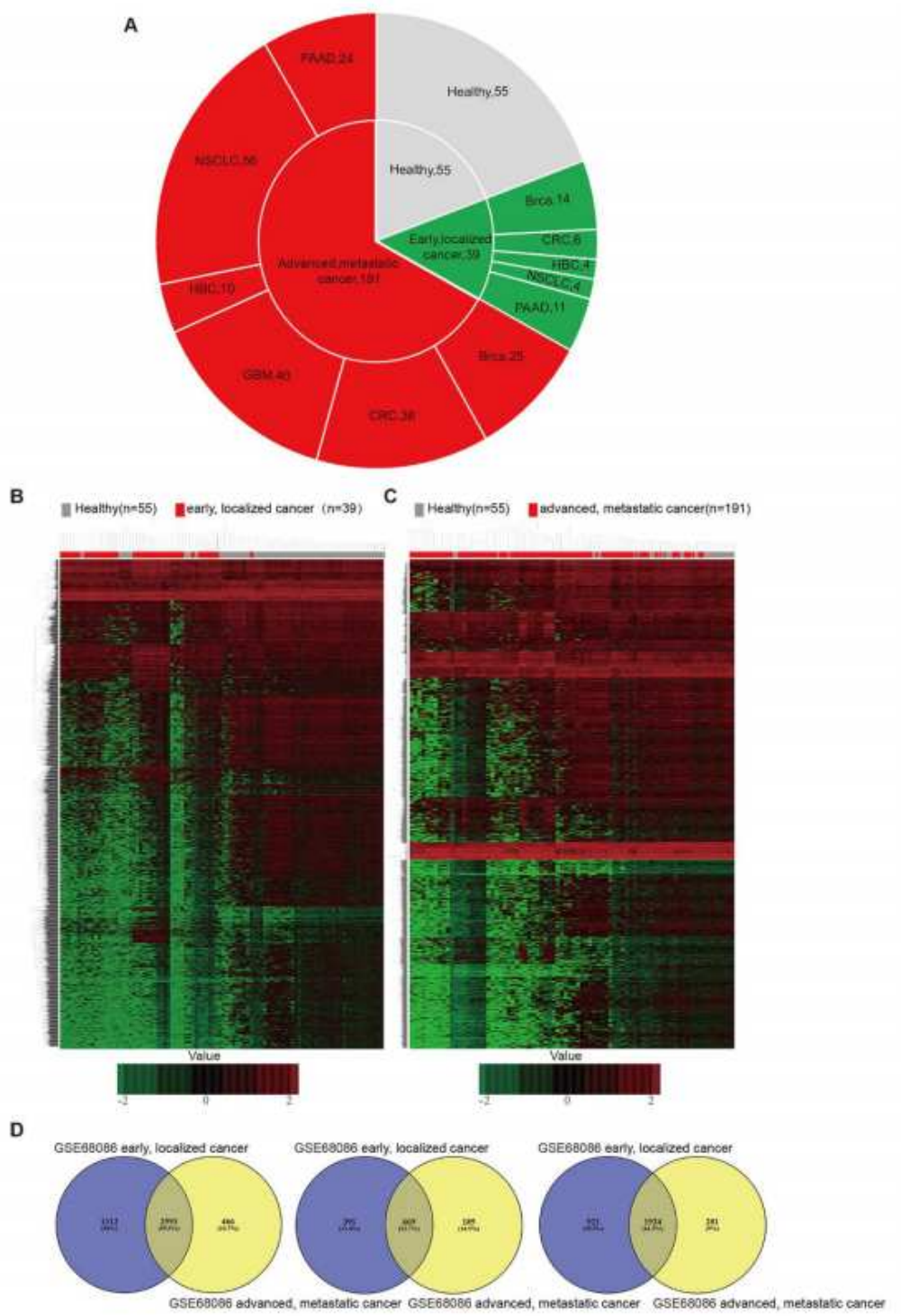

Figure 1

Identification of DEGs in TEPs from localized or metastatic pan cancer patients incomparation with platelets from healthy controls based on the datasets GSE68086. (A), Number of platelet samples of healthy controls and cancer patients with different stages or types of cancer based on GSE68086. (B-C),

Hierarchical clustering heatmap of DEGs in the expression profiling datasets GSE68086. (B), Heatmap of DEGs in TEPs collected from healthy controls and early, localized cancer patients. (C), heatmap of DEGs in TEPs collected from healthy controls and advanced, metastatic cancer patients. The horizontal axis indicates the sample, and the vertical axis indicates the DEGs. Red represents the up-regulated DEGs and green represents the down-regulated DEGs. DEGs, differential expressed genes. (D), Identification of TEPs mRNAs between localized and metastatic pan cancer. Left, commonly altered differential expressed TEPs mRNAs. Middle, identification of up-regulated differential expressed TEPs mRNAs. Right, identification of down-regulated differential expressed TEPs mRNAs

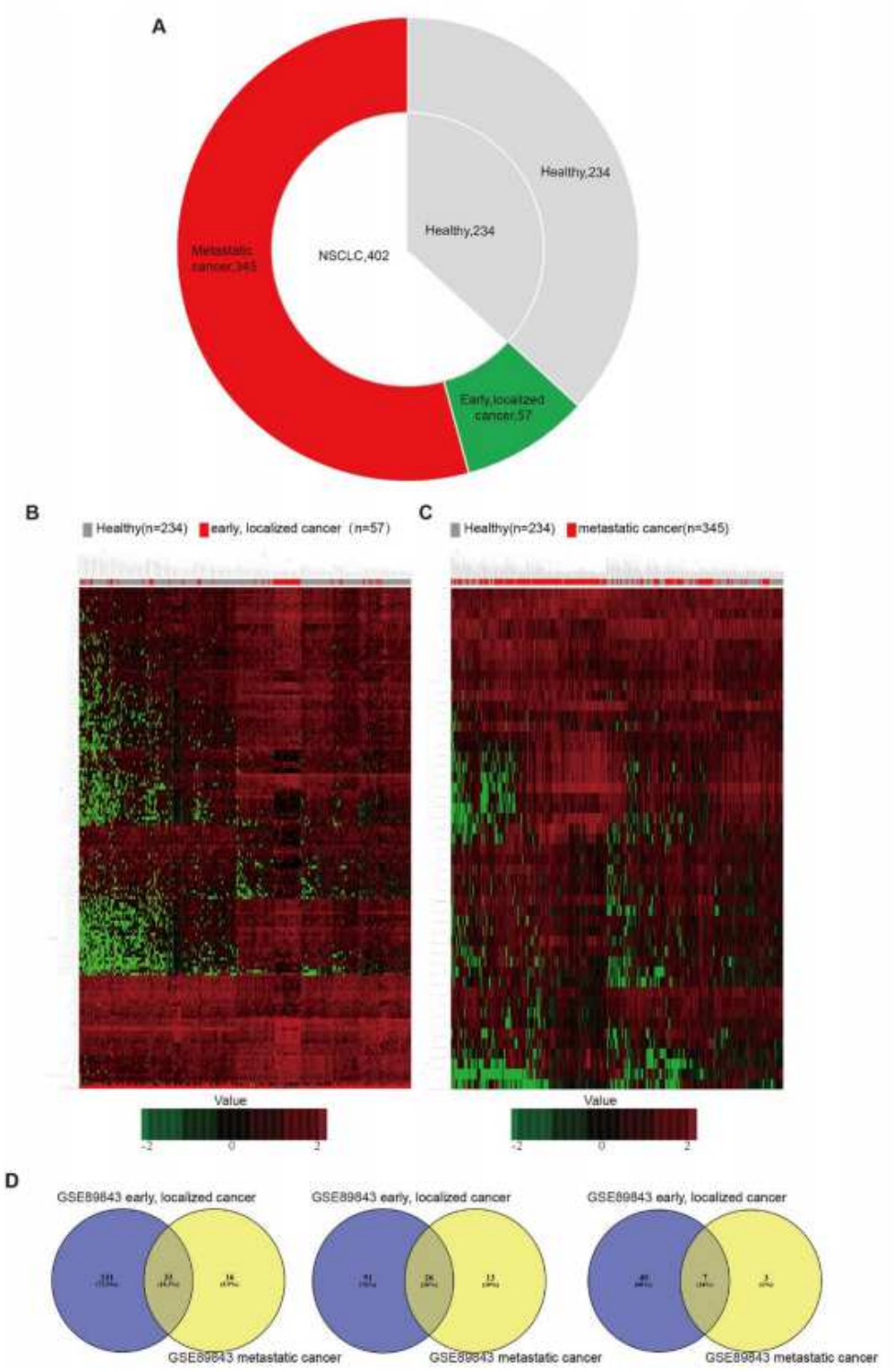

Figure 2

mRNA profiles of TEPs from localized or metastatic NSCLC cancer patients as compared to platelets from healthy controls based on the datasets GSE89843. (A), Number of platelet samples of healthy controls and NSCLC cancer patients at different stages. (B-C), Hierarchical clustering heatmap of DEGs in the expression profiling datasets GSE89843. (B), Heatmap of DEGs in TEPs collected from healthy controls and early, localized NSCLC cancer patients. (C), heatmap of DEGs in TEPs collected from healthy

controls and metastatic NSCLC cancer patients. The horizontal axis indicates the sample, and the vertical axis indicates the DEGs. Red represents the up-regulated DEGs and green represents the 35 / 36 down-regulated DEGs. (D), Identification of TEPs mRNAs between localized and metastatic NSCLC cancer. Left, commonly altered differential expressed TEPs mRNAs. Middle, identification of up-regulated differential expressed TEPs mRNAs. Right, identification of down-regulated differential expressed TEPs mRNAs.

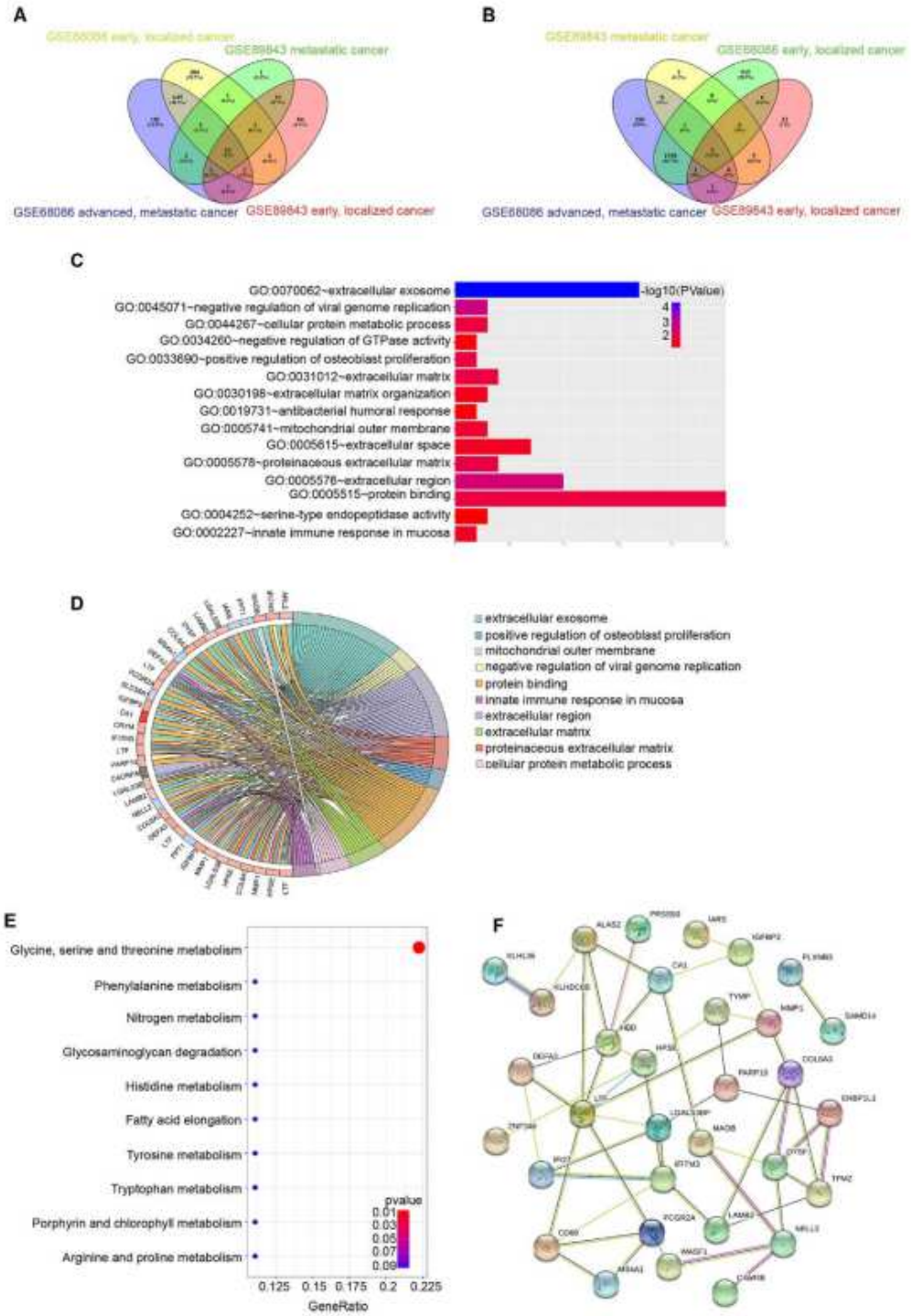

Figure 3

Analysis of the spliced RNA repertoire of TEPs from pan cancer patients at different stages. (A-B), Identification of DEGs in the four datasets (GSE68086 early pan cancer and metastatic pan cancer, GSE89843 early NSCLC cancer and metastatic NSCLC cancer) via Venn diagrams software. (A), identification of up-regulated differential expressed TEPs mRNAs. (B), identification of down-regulated differential expressed TEPs mRNAs. Different colors represent different datasets. (C-D), GO analyses of the DEGs according to their biological process, cellular component and molecular function. GO, gene ontology. (E), KEGG pathway enrichment analysis. Dot size represents the number of genes in each KEGG pathway; P-value: Red < purple < blue. KEGG, Kyoto Encyclopedia of Genes and Genomes. (F) Protein-protein interaction network of DEGs visualized through String datasets.

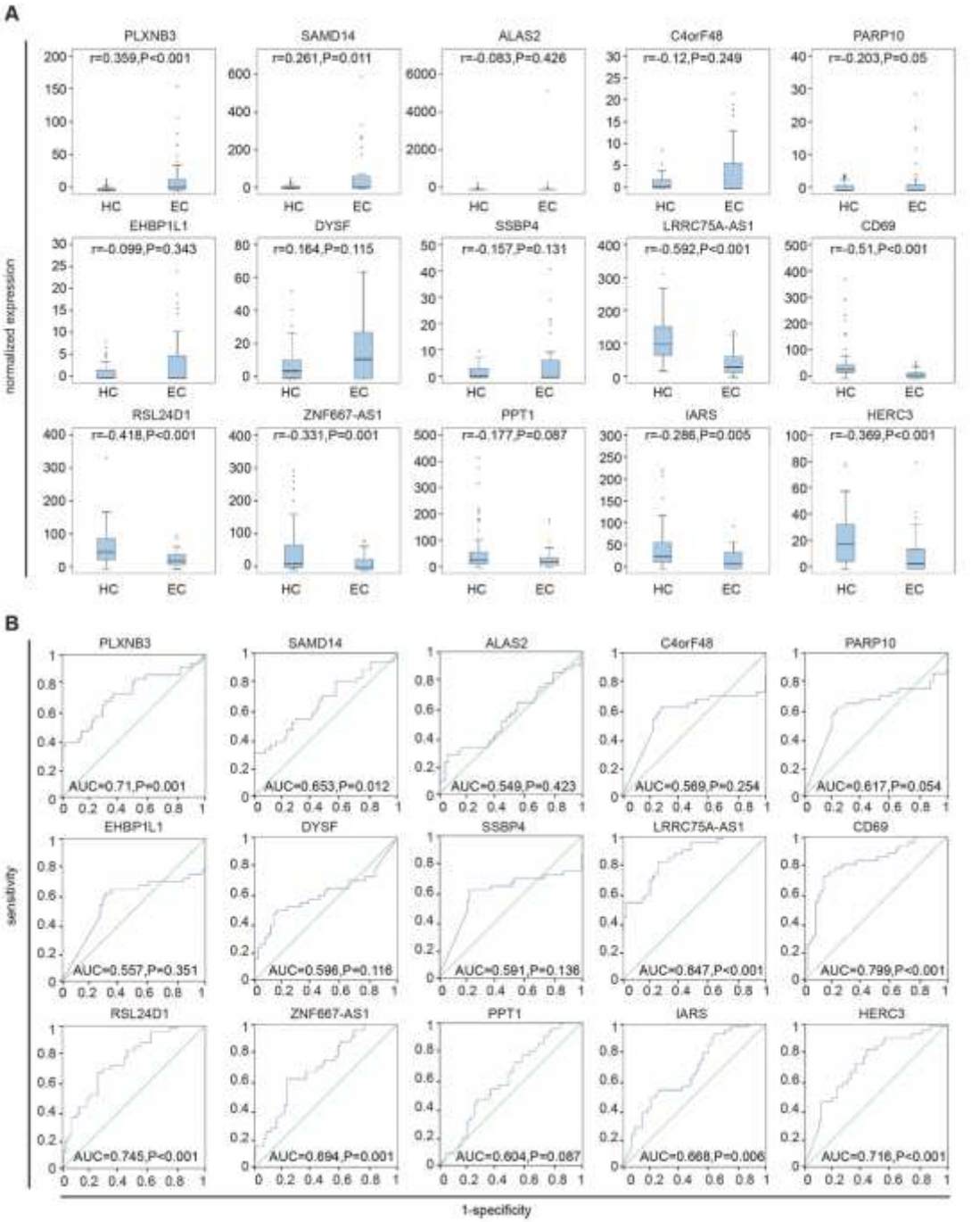

Figure 4

Diagnostic value of TEPs DEGs for early, localized pan cancer based on the datasets GSE68086. (A), Correlation analysis between expression levels of 15 TEPS DEGs and two groups including healthy control groups and early cancer groups. HC, healthy control; EC, early cancer; The 15 TEPS DEGs are PLXNB3, SAMD14, ALAS2, C4orf48, PARP10, EHBP1L1, DYSF, SSBP4, LRRC75A, CD69, RSL24D1, ZNF667, PPT1, IARS, and HERC3, respectively. (B), ROC analysis of sensitivity and specificity of the above 15 TEPS DEGs signature in predicting the diagnosis of early pan cancer patients.

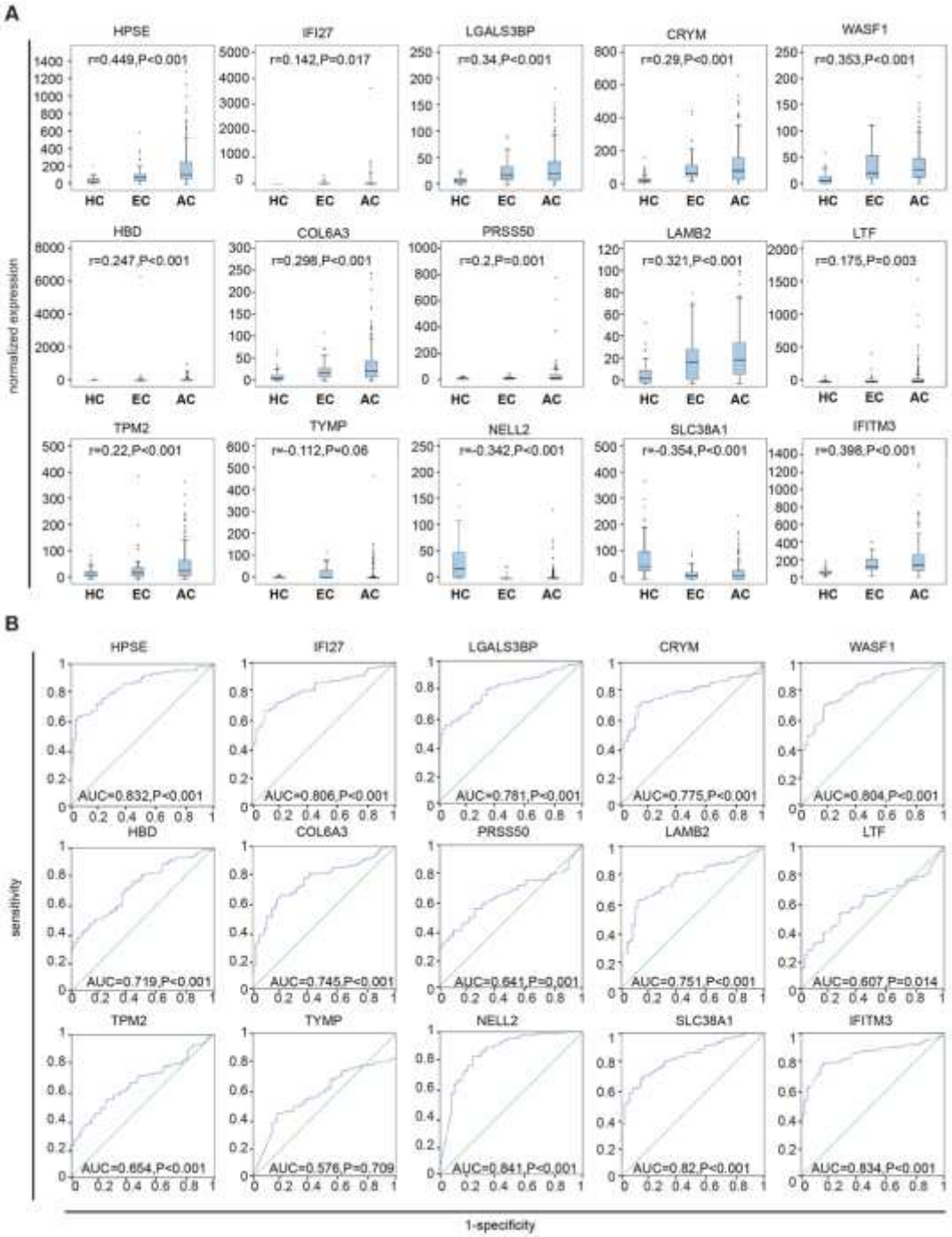

Figure 5

Identification of 15 TEPS DEGs signatures for pan cancer diagnosis based on the 36 / 36 datasets GSE68086. (A), Correlation analysis between expression levels of 15 TEPS DEGs and three groups including healthy control groups, early cancer groups and advanced cancer groups. HC, healthy control; EC, early cancer; AC, advanced cancer. The 15 TEPS DEGs are HPSE, IFI27, LGALS3BP, CRYM, WASF1, HBD, COL6A3, PRSS50, LAMB2, LTF, TPM2, TYMP, NELL2, SLC38A1, and IFITM3, respectively. (B), ROC analysis of sensitivity and specificity of the above 15 TEPS DEGs signature in predicting the diagnosis of pan cancer patients

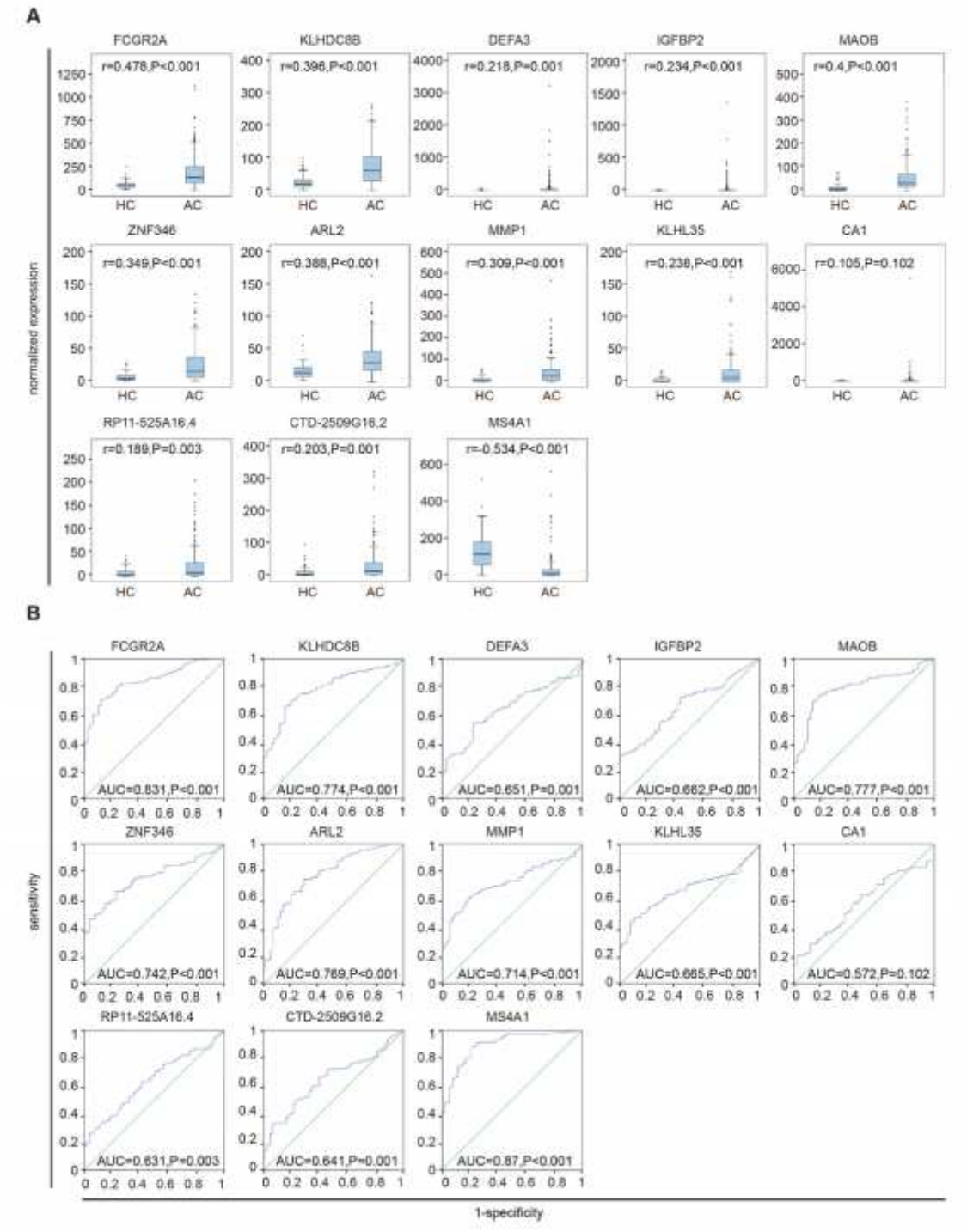

Figure 6

Diagnostic value of 13 TEPs DEGs for advanced, metastatic pan cancer based on the datasets GSE68086. (A), Correlation analysis between expression levels of 13 TEPS DEGs and two groups including healthy control groups and advanced cancer groups. HC, healthy control; AC, advanced cancer. The 13 TEPS DEGs are FCGR2A, KLHDC8B, DEFA3, IGFBP2, MAOB, ZNF346, ARL2, MMP1, KLHL35, CA1, RP11-525A16.4, CTD-2509G16.2, and MS4A1, respectively. (B), ROC analysis of sensitivity and specificity of the above 15 TEPS DEGs signature in predicting the diagnosis of advanced pan cancer patients.

Supplementary Files

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

- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS1.xlsx
- TableS2.xlsx
- supplementalfigures.docx