Molecular Therapy
Nucleic Acids
Original Article

An EV-Associated Gene Signature Correlates with Hypoxic Microenvironment and Predicts Recurrence in Lung Adenocarcinoma

Bangrong Cao,[1,6] Wei Dai,[2,6] Shiqi Ma,[1] Qifeng Wang,[3] Mei Lan,[3] Huaichao Luo,[5] Tingqing Chen,[1] Xiaojun Yang,[2] Guiquan Zhu,[4] Qiang Li,[2] and Jinyi Lang[3]

[1]Radiation Oncology Key Laboratory of Sichuan Province, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China; [2]Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China; [3]Radiation Oncology Key Laboratory of Sichuan Province, Department of Radiation Oncology, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China; [4]Radiation Oncology Key Laboratory of Sichuan Province, Department of Head and Neck Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China; [5]Department of Clinical Laboratory, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

Extracellular vesicles (EVs) mediate intercellular communications in the tumor microenvironment and contribute to the aggressive phenomenon of cancers. Although EVs in body fluids are supposed to be ideal biomarkers for cancer diagnosis and prognosis, it remains difficult to distinguish the tumor-derived EVs from those released by other tissues. We hypothesized that analyzing the EV-related molecules in tumor tissues would help to estimate the prognostic value of tumor-specific EVs. Here, we investigate the expression of coding genes of proteins carried by small EVs (sEVs) in primary lung adenocarcinoma. Based on the protein-protein interaction network, we identified three network modules (3-PPI-Mod) as a signature that could predict recurrence. This signature was validated in three independent datasets and demonstrated better prognostic value than signature generated from gene expression alone. Meanwhile, the high-risk subgroup assigned by the signature could benefit from adjuvant chemotherapy, although it was not beneficial in unselected patients. Two out of three modules were enriched by proteins identified in sEVs from non-small-cell lung cancer cells. Furthermore, the two modules were remarkably correlated with intratumoral hypoxia score. These results suggest that the 3-PPI-Mod signature was enriched in tumor-derived sEVs and could serve as a prognostic and predictive biomarker for lung adenocarcinoma.

INTRODUCTION

Lung cancer is one of the most common malignancies worldwide.[1] According to pathomorphology, nearly 85% of all lung cancers are non-small-cell lung cancers (NSCLCs).[2] Lung adenocarcinoma (ADC) accounts for 35% of NSCLCs. Despite the improvement of treatment, the 5-year overall survival of NSCLC is below 20%.[3] Loco-regional and distant relapse are the major causes of cancer-related death for NSCLC. Even in stage I NSCLC, about 30% of patients would suffer recurrent disease, and many patients would die of it after surgical treatment.[3,4] Although adjuvant chemotherapy (ACT) improves outcome in late-stage NSCLC, the survival benefit of ACT in early-stage patients remains controversial.[5] Thus, exploring novel prognostic biomarkers would be helpful to select patients for ACT and improve prognosis in early-stage NSCLCs.

Extracellular vesicles (EVs) are cell-released membrane particles that regulate intercellular communication by transporting functional molecules (e.g., proteins, nucleic acids, and lipids) to recipient cells.[6] Cells can secret diverse types of EVs with a wide range of sizes and different cellular origins. Exosomes mainly range from 50–150 nm in diameter, which are generated inside multivesicular endosomes.[7] Other EVs (with a diameter from 100 to 1,000 nm), such as microvesicles (MVs), ectosomes, or microparticles, are directly budded from cellular membrane.[7] Given that the specific subcellular origin was not strictly demonstrated in many literatures, the generic term EVs is used in this study, as recommended by the International Society for Extracellular Vesicles (ISEV).[8]

In recent years, increasing evidences have suggested that EVs play an important role in cancer development and progression.

Received 28 September 2018; accepted 28 July 2019;
https://doi.org/10.1016/j.omtn.2019.07.021.

[6]These authors contributed equally to this work.

Correspondence: Jinyi Lang, Radiation Oncology Key Laboratory of Sichuan Province, Department of Radiation Oncology, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, 55 South Renmin Ave. Fourth Section, Chengdu, Sichuan Province, China.
E-mail: langjy610@sichuancancer.org
Correspondence: Qiang Li, Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, 55 South Renmin Ave. Fourth Section, Chengdu, Sichuan Province, China.
E-mail: luckydavii@sina.com

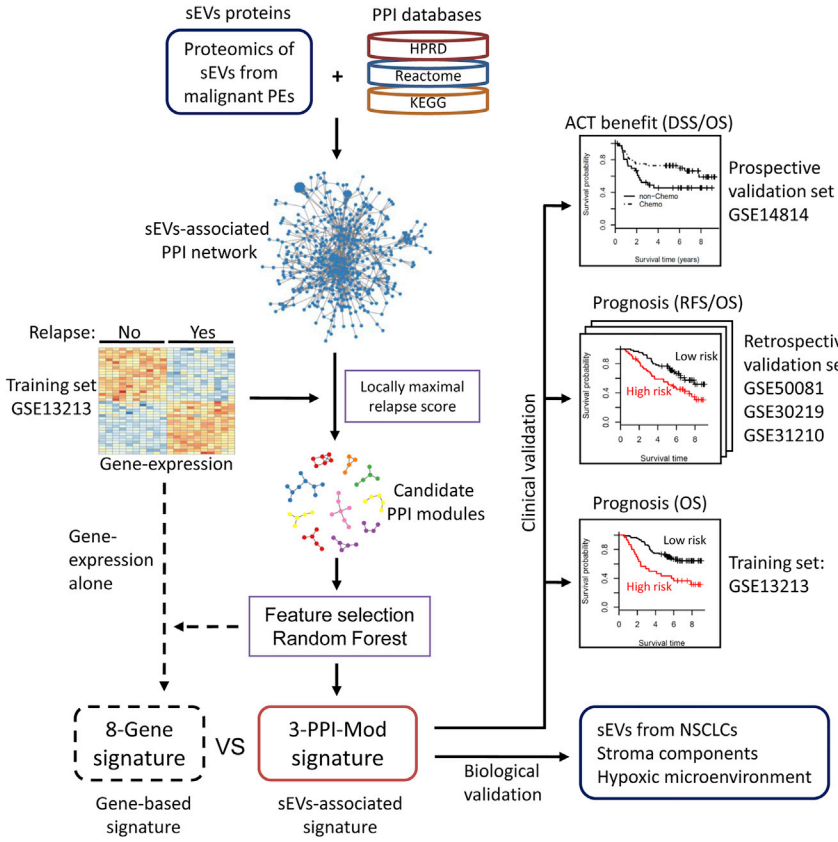

Figure 1. The Workflow of This Study

The PPI network of sEV proteins isolated from PEs was integrated with the gene expression profile of the training dataset. Candidate modules with locally maximal relapse scores were identified by a greedy searching approach. Random forest algorithm was used to establish a network-based signature for relapse risk. The signature 3-PPI-Mod was further validated for prognosis in three independent datasets. Moreover, the predictive value for adjuvant chemotherapy in ADCs was evaluated in a prospective dataset conducted from the JBR.10 trial. Biological validation was performed by comparing the signature with various databases. Abbreviations: PPI, protein-protein interaction; sEVs, small extracellular vesicles; 3-PPI-Mod, 3 PPI modules; ADC, adenocarcinoma; RFS, relapse-free survival; DSS, disease-free survival; OS, overall survival.

Cancer-derived EVs are implicated in various carcinogenesis processes, including malignant transformation, angiogenesis, immunosuppression, invasion, and treatment resistance.[9–11] On the other hand, EVs released by the tumor microenvironment could also influence the traits of cancer cells.[12] Hypoxia, a common feature of tumor microenvironment, was demonstrated to induce cancer EVs that transport aggressive and metastatic phenomenon to the recipient cells.[13–16] Moreover, cancer-released EVs could modify the distant microenvironment to form a pre-metastatic niche to facilitate the formation of metastatic lesions, suggesting that cancer EVs could act as both local and distant effects.[17]

Cancer-derived EVs detected in various body fluids are proposed as novel noninvasive biomarkers. For instance, circulating small EVs (sEVs, with a diameter less than 200 nm) carried microRNAs (miRNAs)[18,19] and proteins,[20–22] which are promising diagnostic and prognostic biomarkers for NSCLC. Park et al.[23] examined the proteomics of sEVs isolated in malignant pleural effusions (PEs) from metastatic lung ADCs. The study identified a lot of well-known cancer-related proteins, such as EGFR, RAS, and Src-family kinase.[23] However, the sEVs identified in body fluids may represent a mixed population released by both tumors and other tissues. To date, it remains a great challenge to distinguish tumor-specific sEVs from those released by other tissues due to the lack of specific biomarkers.[24] It was reported that the protein levels in sEVs released to body fluids was consistent with protein expression of the NSCLC tissues.[25] Therefore, we proposed that integrative analysis of primary tumors with secreted sEVs may be feasible to identify tumor-specific sEV-related biomarkers for NSCLC.

In this study, we analyzed the intratumoral gene-expression profile of the coding genes of sEV proteins identified in malignant PEs from advanced lung ADC.[23] Based on a protein-protein interaction (PPI) network approach, a sEV-associated gene signature was established for recurrence prediction. This signature was further validated for its prognostic value in three retrospective datasets and one prospective dataset from the public database (Table S1). Moreover, the gene signature was further compared with proteomics of sEVs isolated from conditional medium of NSCLC cells as well as peripheral blood from NSCLC patients. The association between the signature and tumor microenvironment, including hypoxic status and stromal cell abundance, were also investigated.

RESULTS

Identifying Relapse-Associated PPI Modules from sEV Proteins

Figure 1 depicts an overall flow chart of this study. For the sEV proteins identified in malignant PEs, 568 proteins were mapped onto the reference PPI network. The sEV-associated PPI network was integrated with the gene-expression profile of the training dataset. Next, the module searching procedure identified 144 modules, which showed locally maximal discriminative scores for recurrent status. Three permutation methods were used to estimate the significance of the discriminative score for each module. As a result, 21 modules showed significantly higher discriminative scores for recurrence than those by chance ($p < 0.001$ for all three methods, Figure S1). Details of 21 candidate modules are presented in Figure S2.

The expression score profiles of the 21 PPI modules are presented in Figure 2A. All modules were clustered into two major clusters by

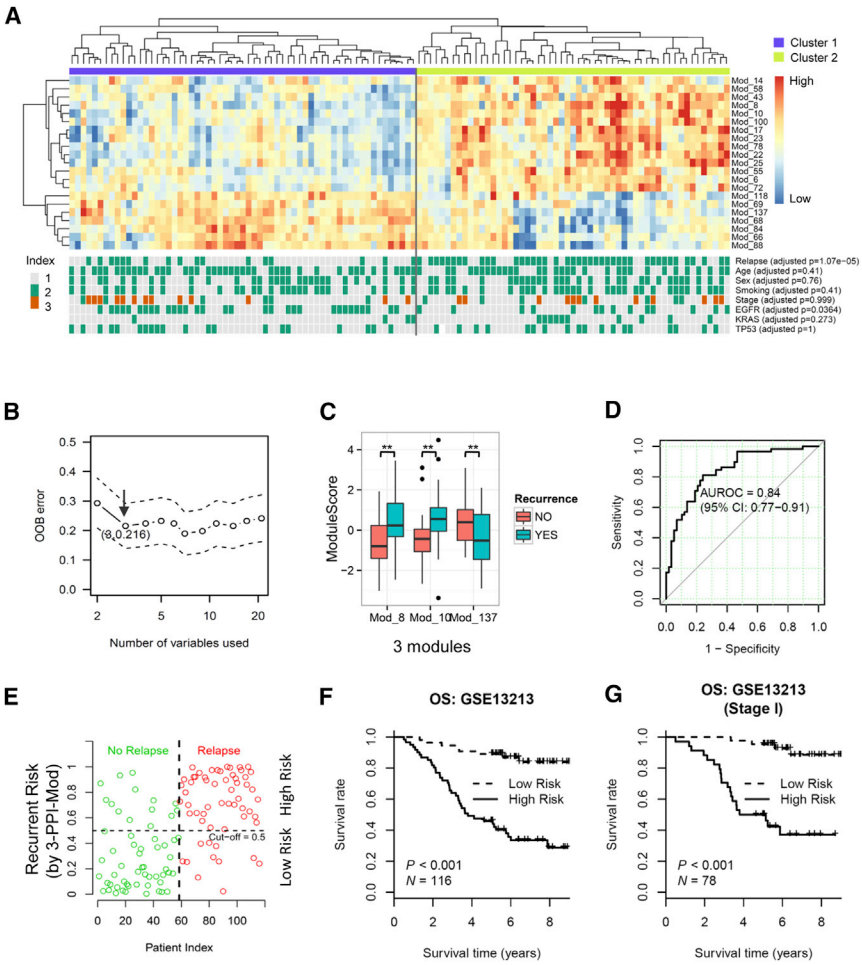

Figure 2. Establishing a Prognostic Signature Based on a sEV-Associated Network

(A) Expression score profiles of candidate PPI modules. The upper heatmap shows the expression scores of the 21 PPI modules. Unsupervised hierarchical clustering analysis was performed for modules (rows) and samples (columns). The lower panel indicates clinicopathological variables of the tumor samples. The p value was calculated by chi-square test and then adjusted by FDR for multiple testing. (B) Feature selection based on RF algorithm. The x axis indicates the number of used variables (modules). The y axis is the out-of-bag (OOB) error of each prediction model. The arrow indicates that the combination of 3 PPI modules achieved the selection criteria with the smallest feature number and maximal prediction accuracy. (C) Expression scores of the 3 modules that were selected in (B). The p value was calculated by t test. (D) The recurrent risk predicted by 3-PPI-Mod was used to perform ROC analysis. AUROC and 95% CI are presented. (E) Scatterplotting of the predictive risk of recurrence. The x axis indicates patient index. The y axis indicates the predictive risk of recurrence (OOB). A cutoff of 0.5 was used to classify patients into low- and high-risk subgroups. Red and green colors indicate patients with and without recurrent disease, respectively. (F and G) Survival analysis for 3-PPI-Mod in all patients (F) and in stage I only (G). Kaplan-Meier curves and log-rank test were used to compare OS between different risk subgroups identified in (E). Abbreviations: PPI, protein-protein interaction; ADC, adenocarcinoma; FDR, false discovery rate; RF, random forest; ROC, receiver operating characteristic curve; AUROC, area under the receiver operating characteristic curve; CI, confident interval; OS, overall survival.

unsupervised clustering algorithm. One cluster with 14 modules was more likely upregulated in recurrent tumors, and another cluster with 7 modules was inversely correlated with recurrence. Patients were also grouped into two major subgroups, which were significantly associated with recurrent status (adjusted $p < 0.001$) and EGFR mutation (adjusted $p = 0.036$) (Figure 2A). However, there was no statistically significant association of patient subgroups with other clinicopathological factors, such as age, sex, smoking index, tumor stage, KRAS status, or TP53 status.

Establishing a Predictive Signature for Cancer Recurrent Risk

We then thought to establish an optimal model to predict recurrent risk for patients using the 21 PPI models. By using the random forest (RF) algorithm, we found that the combination of three PPI modules achieved the criteria of the smallest feature number and the best prediction accuracy (Figure 2B, out-of-bag [OOB] error = 0.216 ± 0.038). Two out of three modules (Mod_8 and Mod_10) were upregulated, and the last module (Mod_137) was downregulated in recurrent tumors (Figure 2C). Next, the three PPI modules were used to construct a prediction model, which was named 3-PPI-Mod signature. The recurrent risk of patients was predicted by the signature, with an area under the receiver operating characteristic curve (AUROC) of 0.84 (95% confident interval [CI], 0.77–0.91) (Figure 2D). Using 0.5 as the cut-off, patients were divided into two groups with high or low risk for recurrence, respectively (Figure 2E). Overall, 78.4% (91/116) of patients were correctly classified when compared with the true recurrent status. Kaplan-Meier curves confirmed that the 3-PPI-Mod signature could predict overall survival (OS) for all patients and those with stage I disease (both $p < 0.001$, Figures 2F and 2G). Multivariate Cox regression demonstrated that the 3-PPI-Mod signature was an independent prognostic factor for OS (adjusted hazard ratio [HR] = 6.58, 95% CI, 3–14.4, $p < 0.001$, Table S2).

Within the 3-PPI-Mod signature, 5, 6, and 5 genes were included in Mod_8, Mod_10, and Mod_137, respectively. One gene (Cofilin-1 [CFL1]) was overlapped by Mod_8 and Mod_10. Mod_10 was connected with Mod_8 and Mod_137 within the PPI network (Figure 3A). Gene Ontology (GO) and pathway enrichment analysis showed that all three modules were significantly correlated with focal

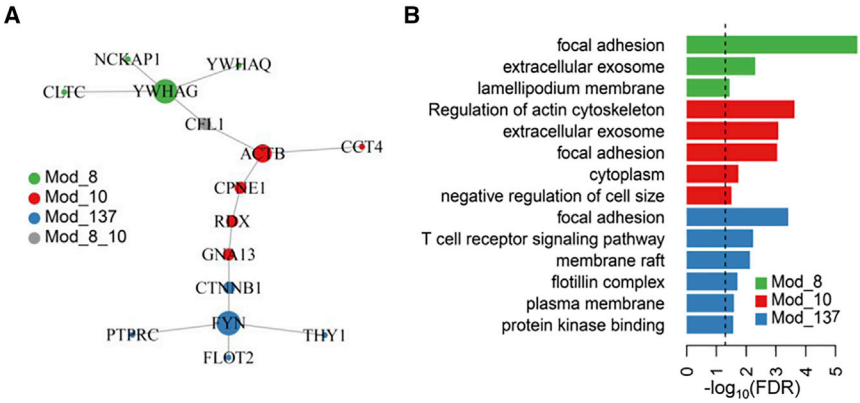

Figure 3. Biological Characteristics of the 3-PPI-Mod Signature

(A) The view of interaction of 3-PPI-Mod signature. The presented network was generated by mapping the signature proteins on the reference PPI network. Three modules are shown by different colors. (B) Enrichment analysis for three modules. The significant GO terms or KEGG pathways were listed on the left. Bars indicate the enrichment significance (log10-transformed FDR). The color legend indicates different modules. Abbreviations: PPI, protein-protein interaction; 3-PPI-Mod, 3 PPI modules; GO, gene ontology; FDR, false discovery rate.

adhesion (Figure 3B). Notably, genes in Mod_8 and Mod_10 were enriched in extracellular exosome (Figure 3B). Mod_137 showed association with diverse biological processes, including protein kinase binding, membrane raft, and T cell receptor signaling pathway (Figure 3B).

Validation of the 3-PPI-Mod Signature in Independent Cohorts

For each validation dataset, the recurrent risks of patients were predicted by the 3-PPI-Mod signature, which was established in the training dataset. Kaplan-Meier curves demonstrated that patients who were classified as high risk had significantly shorter relapse-free survival (RFS) time than those classified as low risk (log-rank test: Gene Expression Omnibus [GEO]: GSE50081, $p = 0.005$; GEO: GSE30219, $p = 0.003$; GEO: GSE31210, $p < 0.001$) (Figures 4A–4C). The adjusted HRs (95% CI) for high-risk versus low-risk subgroups were 2.24 (1.07–4.67), 3.74 (1.54–9.07), and 3.34 (1.81–6.15) in GEO: GSE50081, GSE30219, and GSE31210, respectively (Table S3). Similarly, the high-risk subgroup was significantly correlated with unfavorable OS compared with the low-risk subgroup (GEO: GSE50081, HR [95% CI] = 1.92 [1.03–3.58], $p < 0.001$; GEO: GSE30219, HR [95% CI] = 2.41[1.25–4.63], $p = 0.006$; GEO: GSE31210, HR [95% CI] = 2.46[1.09–5.58], $p = 0.001$) (Figures 4D–4F; Table S3). Subgroup analysis revealed that the 3-PPI-Mod signature remained significant in distinguishing RFS and OS for patients with stage I disease (Figures S3A–S3F).

We further validated the 3-PPI-Mod signature in a prospective dataset conducted from the JBR.10 clinical trial, which was designed to evaluate the advantage of ACT in stage IB to II NSCLCs. When all ADC patients were taken into consideration, the ACT arm showed no significant improvement of disease-specific survival (DSS) compared with the observation (OBS) arm (Figure 5A, log-rank test, $p = 0.503$). The 3-PPI-Mod signature classified 32 and 39 of 71 ADC patients as low- and high-risk subgroups, respectively. Although ACT did not improve outcome in the low-risk subgroup (Figure 5B, log-rank test, $p = 0.189$), it significantly prolonged DSS in the high-risk subgroup (Figure 5C, log-rank test, $p = 0.019$). Similar results were observed in OS analysis (ACT versus OBS, log-rank test: all ADCs, $p = 0.907$; low-risk subgroup, $p = 0.115$; high-risk subgroup, $p = 0.064$) (Figures 5D–5F). Multivariate Cox's model revealed a significant interaction effect between risk groups and ACT (test for interaction: DSS, $p = 0.011$; OS, $p = 0.013$) (Table S4). For non-ADC patients, however, there was no evidence demonstrating that ACT could improve DSS or OS, even if in subgroups stratified by the 3-PPI-Mod signature (Figure S4; Table S4).

The sEV-Associated Signature Exhibited Better Prognostic Value Than Signature Derived from Gene Expression Alone

The sEV-associated 3-PPI-Mod signature was compared with an 8-gene signature, which was constructed based on gene expression alone (Figures S5A–S5D). The 8-gene signature was established by using the same algorithm as used for the 3-PPI-Mod signature. In the training dataset, the 8-gene signature showed similar prediction performance as the 3-PPI-Mod signature in distinguishing relapse and OS (Figures S5E and S5F). In the validation datasets, however, the C-index of 3-PPI-Mod signature was higher than that of the 8-gene signature in predicting both RFS and OS (Figures S5G and S5H). Kaplan-Meier analysis also confirmed that the 8-gene signature could not consistently distinguish RFS and OS across different validation cohorts when considering neither all patients nor the stage I subgroup (Figures S6A and S6B). Meanwhile, the 8-gene signature could not predict ACT benefit in the JBR.10 dataset (Figure S7). These findings suggest that the sEV-associated signature exhibited improved prognostic value compared with signature derived from gene expression alone, especially for independent cohorts.

The 3-PPI-Mod Reflects Tumor-Stroma Interaction and Hypoxic Tumor Microenvironment

The proteins within the 3-PPI-Mod signature were further validated by the proteomics of sEVs isolated from NSCLC cell supernatants[26] and serums of NSCLC patients.[22] Overall, 10 and 8 proteins of the signature were overlapped with those identified in sEVs from NSCLC cells and circulating serums, respectively (both $p < 0.001$, by hypergeometric test) (Figures 6A and 6B). More than 90% of the overlapped proteins came from Mod_8 and Mod_10. For Mod_137, only one protein (CTNNB1) was overlapped with proteomics of

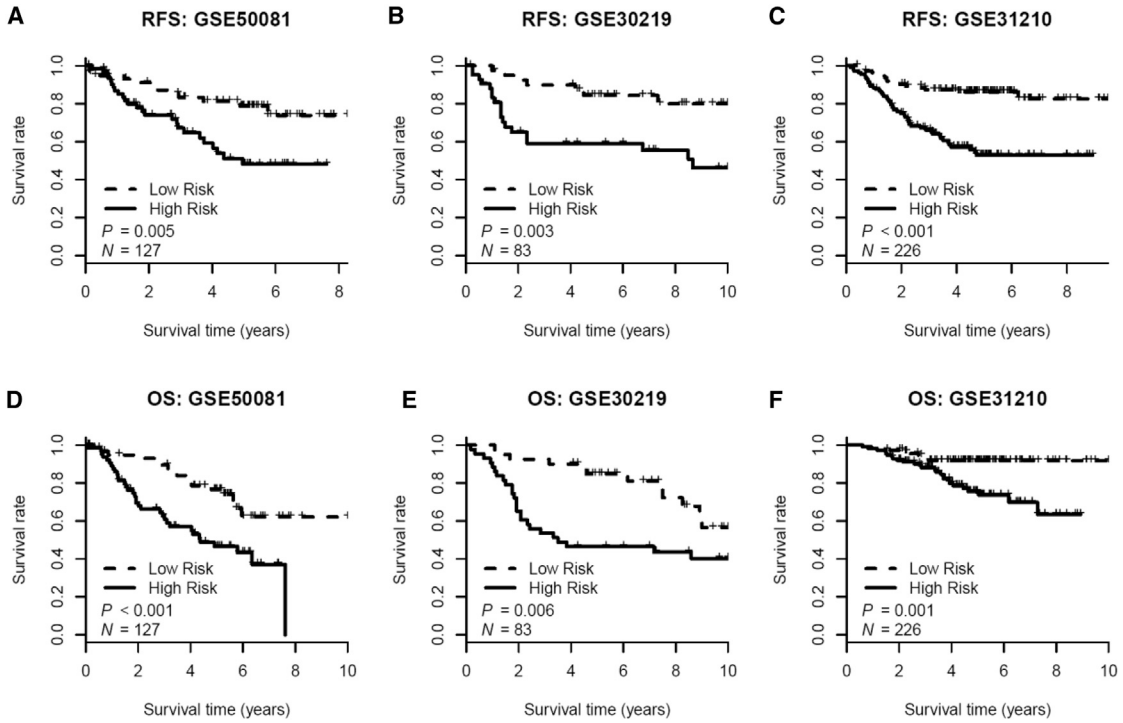

Figure 4. Survival Analysis of 3-PPI-Mod Signature in Three Validation Datasets

(A–F) In each dataset, patients were predicted to be in low- or high-risk subgroups by the signature. Kaplan-Meier curves and log-rank test were used to compare RFS (A–C) and OS (D–F) of patients in different risk subgroups: GEO: GSE50081 (A and D), GSE30219 (B and E), GSE31210 (C and F). Survival data within 10 years are shown. Abbreviations: 3-PPI-Mod, 3 PPI modules; RFS, relapse-free survival; OS, overall survival.

NSCLC-derived sEVs and none with that of circulating sEVs (Figures 6A and 6B).

We hypothesized that Mod_137 may be associated with non-cancer cells. The abundance of different stromal cells was estimated by microenvironment cell population (MCP)-counter.[27] Surprisingly, a higher expression score of Mod_137 was correlated with elevated abundance of fibroblasts, endothelial cells, and diverse types of immune cells (Figure 6C). In contrast, Mod_8 and Mod_10 expression scores did not show a remarkably positive correlation with those of the stromal components. Similar results were observed when analysis was conducted by using individual genes within each module (Figure 6C).

The association of 3-PPI-Mod signature with tumor microenvironment status was analyzed. We found that Mod_8 and Mod_10 were positively correlated with hypoxia, while Mod_137 was associated with angiogenesis and inflammation (Figure 6C). The hypoxia metagene was significantly enriched in high-risk tumors predicted by the 3-PPI-Mod signature (new enrichment score [NES] = 1.88, q value = 0.005) (Figure 6D). There was a significant correlation between intratumoral hypoxic score and hypoxia inducible factor 1 (HIF1) expression level, which were both upregulated in the high-risk subgroup identified by the 3-PPI-Mod signature (Figures 6E–6G). However, the angiogenesis and inflammation status was not associated with 3-PPI-Mod subgroups, based on gene set enrichment analysis (GSEA) (Figures S8A and S8B) or expression scores (Figures S8C and S8D).

DISCUSSION

In the present study, we evaluate the intratumoral gene expression level of coding genes of sEV proteins in lung ADC. In the training dataset, we identified a 3-PPI-Mod signature that could distinguish patients with high risk of recurrence. This signature was also validated in three independent datasets and demonstrated better prognostic value than the gene signature established by using gene expression alone. Moreover, the patients assigned as high risk by the signature could benefit from ACT, although no significant benefit was achieved for unselected patients. Further analysis revealed that the sEV-associated signature is highly correlated with intratumoral hypoxic status and might be derived from both cancer cells and stromal cells. To the best of our knowledge, this is the first comprehensive analysis of sEV-associated genes in predicting outcome of lung ADC based on large-number populations from several independent cohorts.

To date, it remains a great challenge to discriminate exosomes from other sEVs such as MVs.[24] Exosomes are derived from endosomal multivesicular bodies that fused with the cytoplasma membrane, while MVs are usually considered to bud directly from the cytoplasma membrane.[9] The PE-derived "MVs"[23] and NSCLC cell-released

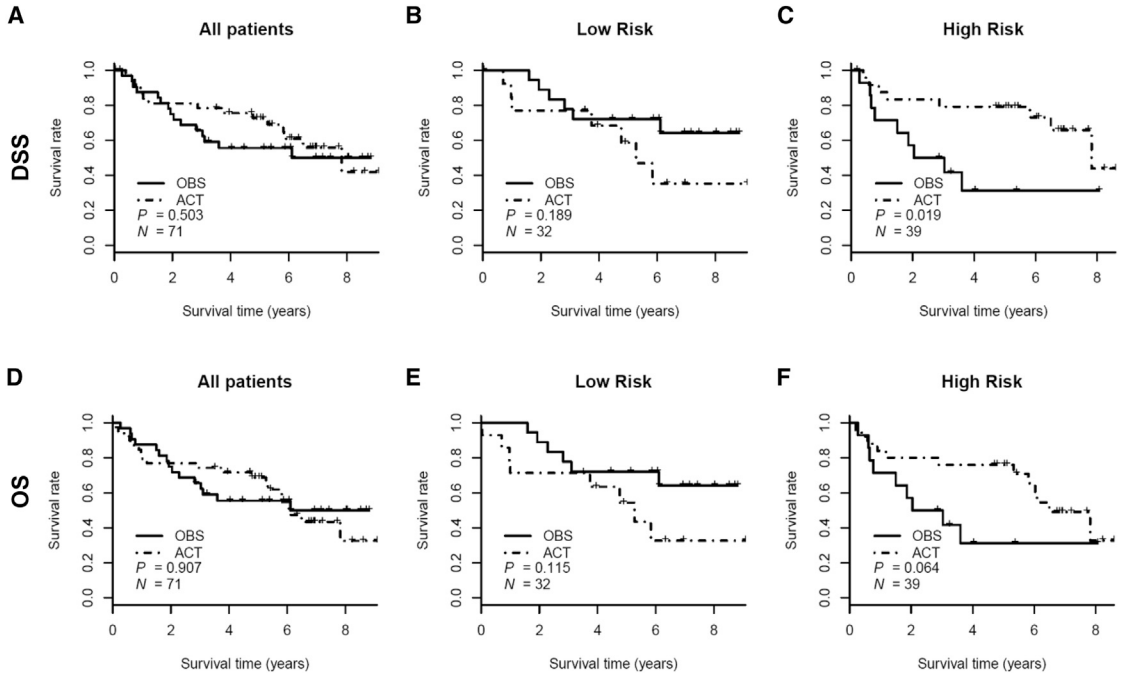

Figure 5. Predictive Value of 3-PPI-Mod Signature for ACT Benefit in ADC Patients from the JBR.10 Dataset

(A–F) Results with lung ADC patients are presented. Postoperative DSS (A–C) and OS (D–F) between ACT arm and observation arm (OBS) were compared by Kaplan-Meier curves and log-rank test. Patients were stratified by 3-PPI-Mod signature. The survival comparisons were performed in all patients (A and D), low risk subgroup (B and E), and high risk subgroup (C and F), respectively. Abbreviations: ACT, adjuvant chemotherapy; 3-PPI-Mod, 3 PPI modules; DSS, disease-specific survival; OS, overall survival; OBS, observation.

"exosomes"[26] were isolated by using differential ultracentrifuge followed by density gradient methodology, which recovers mixed EV populations. For these EVs, the multivesicular body origin or purity were not further demonstrated. Therefore, we used the general term sEVs in this study to refer to these EVs, which were characterized with diameters less than 200 nm.[23,26]

The sEV cargos are suggested as promising cancer biomarkers for non-invasive detection due to the presence in various body fluids (e.g., plasma, PEs, sputum, and urine). However, it remains difficult to evaluate the role of tumor-specific sEVs, which may be influenced by those released from other tissues.[24] It was reported that the protein expression of sEVs released to body fluids was consistent with that of the tumor tissues, suggesting feasibility for analyzing primary tumors to reflect the status of circulating sEVs.[25] In this study, by investigating the intratumoral expression of coding genes of proteins carried within sEVs,[23] we identified a prognostic gene signature for lung ADC. Most proteins of the signature were confirmed by sEV proteomics from NSCLC cells,[26] suggesting that the signature may be enriched by tumor-specific sEVs. It is possible that the gene expression in tumor tissues could influence cancer progression. Nevertheless, secretion of the gene products to circulating blood through sEVs may further promote recurrence or distant metastasis.[17,28] This may be explainable by our observation that the sEV-associated signature exhibited better prognostic value than signature that was established from gene expression alone. Therefore, this study provides a promising strategy to evaluate the prognostic value of tumor-specific sEVs from complex body fluids.

The human PPI network has been demonstrated to be informative in biomarker development for cancer when being integrated with gene-expression profiles.[29] Based on the PPI network approach, we identified three network modules that were highly correlated with cancer recurrence. Although these modules were established in the training dataset, their prognostic value was confirmed in external cohorts with hundreds of patients. As we previously reported, the PPI network approach could identify vital biological themes that reflect mechanisms underlying cancer behaviors.[30] In this study, most of the modules (e.g., Mod_8 and Mod_10) were associated with exosome, focal adhesion, and cytoskeleton regulation, which are involved in cancer invasion and metastasis.[31] These results suggest that the 3-PPI-Mod signature represents malignant traits of cancer and is robust in outcome prediction across different cohorts.

Hypoxia, a frequent feature in solid tumors, could induce lung cancer cells to release more sEVs that promote migration via miR-23a delivery.[16] In our signature, Mod_8 and Mod_10 were positively correlated with a hypoxia score that could predict prognosis of NSCLC as reported previously.[32] This was consistent with our findings that Mod_8 and Mod_10 were associated with higher recurrent risk. CFL1, a common

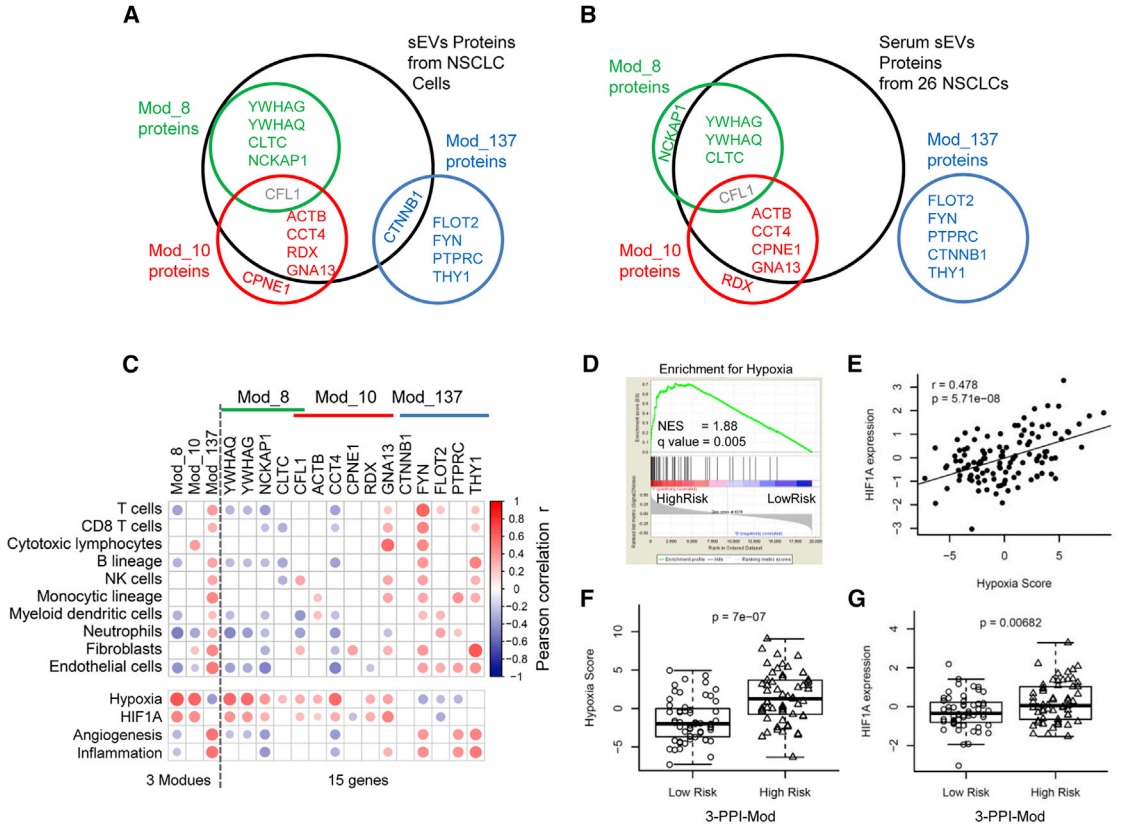

Figure 6. The 3-PPI-Mod Reflects Interaction between Cancer Cells and Tumor Microenvironment

(A and B) Comparison of 3-PPI-Mod signature with other proteomic databases of sEVs from NSCLCs. The venn diagram shows the overlapping of proteins of three modules with those identified in sEVs from NSCLC cell supernatants (A) and circulating serum (B). (C) Correlation analysis of the signature with stromal cell abundance and microenvironment status. The abundance of 8 immune cells and 2 non-immune cells were calculated by MCP-counter. The microenvironment status for hypoxia, angiogenesis, and inflammation were estimated as described in Materials and Methods. Circles in the matrix indicate the correlation coefficient *r*, and red and blue colors indicate positive and negative correlations, respectively. A larger circle size represents a higher correlation. The raw *p* value was adjusted by FDR. Correlations with FDR > 0.05 were not presented. (D) GSEA of hypoxia metagene against the risk subgroups classified by 3-PPI-Mod. The NES and *q* value (adjusted *p* value) are presented. (E) Diagram shows Pearson's correlation between hypoxic score and HIF1A expression level. (F) Boxplot shows intratumoral hypoxic scores between low- and high-risk subgroups. The *p* value was calculated by *t* test. (G) Boxplot shows HIF1A expression levels between tumors with low and high risk. The *p* value was calculated by *t* test. Abbreviations: PPI, protein-protein interaction; 3-PPI-Mod, 3 PPI modules; FDR, false discovery rate; NSCLC, non-small cell lung cancer; MCP, microenvironment cell populations; GSEA, gene set enrichment analysis; NES, new enrichment score.

gene between Mod_8 and Mod_10, was correlated with invasiveness and unfavorable outcome in NSCLC.[33,34] Functional analysis revealed that CFL1 participates in actin dynamics and contributes to membrane extension direction in cell migration.[35] Moreover, many other genes (WYHAG, NCKAP1, ACTB, GNA13, and RDX) identified in Mod_8 and Mod_10 also play an important role in cancer invasion and metastasis.[36–40] However, the association with hypoxia of these genes in NSCLC remains unclear. We found that most of these genes were positively correlated with intratumoral hypoxic score. Thus, we propose that the expression and prognostic role of the sEV-associated signature may be promoted by hypoxic microenvironment, although extended experimental validation is required.

Unlike Mod_8 and Mod_10, Mod_137 had only one protein that was presented in sEVs from NSCLC cells. By analyzing the microenvironment cell populations,[27] we found that four proteins in Mod_137 were correlated with diverse stromal cells. This observation was supported by the fact that FYN, PTPRC (CD45), and THY1 (CD90) might be involved in lymphocytes,[41,42] macrophages,[43] and fibroblasts[44] in tumor microenvironment. Moreover, PTPRC was also identified in sEVs released by dentritic cells.[45] Interestingly, four proteins in Mod_137 were significantly correlated with endothelial cell abundance, which was consistent with the fact that Mod_137 was associated with angiogenesis score. Overall, the sEV proteins derived from cancer cells together with stromal cells may contribute to aggressive microenvironment and influence postoperative recurrence in lung ADC.

One limitation of this study is the relative small sample set of the sEVs cohort, which was used to establish the gene signature. To address this

issue, the 3-PPI-Mod signature was further compared with proteomics of circulating sEVs from another cohort with 26 NSCLC patients.[22] The results confirmed that most proteins of Mod_8 and Mod_10 were presented in circulating sEVs, which was isolated by anti-CD9 antibodies.[22] Notably, none of the proteins in Mod_137 were identified in these circulating CD9+ sEVs. This could be explained by the fact that Mod_137 is probably a representative for sEVs from non-cancer cells. Meanwhile, it was reported that a subpopulation of major histocompatibility complex (MHC) class II-bearing sEVs from dentritic cells could escape from anti-CD9 capture.[45] Taken together, these findings suggest that the 3-PPI-Mod signature is consistently present in different sEV proteomic datasets of NSCLC.

In summary, this study developed a sEV-associated gene-expression signature (3-PPI-Mod) that could predict prognosis and guide ACT benefit for lung ADC. The signature was enriched by cancer-secreted sEVs and was remarkably associated with intratumoral hypoxic status. This study provided a promising strategy to evaluate the prognostic value of sEV-related molecules in tumor tissues and shed new light on the mechanisms underlying metastasis and recurrence of lung ADC.

MATERIALS AND METHODS

Patients and Clinical Characteristics

The proteomic database of sEVs isolated from malignant PEs from three patients with lung ADC was obtained from Park et al.'s [23] study. There were two females and one male. All of the patients suffered stage IV disease, and two patients had metastases at distant organs. Another proteomic database of serum sEVs from 26 NSCLC patients was employed for validation, including 14 cases with lung ADC.[22]

Gene expression profiles of primary lung ADCs from 623 patients were collected from five independent datasets in GEO. These datasets were characterized with geographical and technical diversity (Table S1). GEO: GSE14814 was conducted from the JBR.10 trial, which included NSCLC patients who were randomly assigned into the ACT and OBS arms. For datasets with multiple pathological types of NSCLC, we analyzed the subgroup of lung ADC only. The clinical parameters and follow-up information of each cohort were obtained from GEO. Cases without prognostic information were excluded from analysis. Finally, there were 116 lung ADC samples in GEO: GSE13213, 127 lung ADC samples in GEO: GSE50081, 83 lung ADC samples in GEO: GSE30219, 226 lung ADC samples in GEO: GSE31210, and 71 lung ADC samples in GEO: GSE14814. Clinical characteristics of patients from each dataset are summarized in Table 1. The involvement of human subjects from publicly available resources is in accordance with the Declaration of Helsinki.

Proteomic Database of sEVs

The proteomics of sEVs isolated from malignant PEs,[23] serum samples,[22] and supernatant of NSCLC cells HCC827 and A549[26] were profiled by mass spectrometer. These proteins were mapped to a PPI network and gene expression profiles by Entrez gene IDs.

Gene-Expression Data Processing

Raw data from the training dataset (GEO: GSE13213) with the Agilent platform were processed with "median-scaled" normalization by R package "limma". Raw data fom four validation datasets (GEO: GSE50081, GSE30219, GSE31210, and GSE14814) with the Affymetrix platform were downloaded from GEO. These data were processed and normalized with "RMA" method by R package "affy". Probe annotation of each dataset was downloaded from the GEO database. All gene expression profiles were $\log_2$-transformed.

Processing of PPI Databases

Three public databases were employed to construct the reference PPI network in this study. PPI network from the Human Protein Reference Database (HPRD, release 9.0)[46] was composed of 9,466 proteins and 36,895 interactions. Protein-protein relationships labeled with "Compound/Binding" in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were extracted, resulting in 2,986 proteins and 31,047 interactions. Protein-protein relationships labeled with "Direct Complex" in Reactome pathways were extracted (4,885 proteins and 22,763 interactions). These databases were merged into a reference PPI network, with 11,591 proteins and 77,059 interactions (after removing duplicated interactions and protein-self loops). Network mapping across different databases was done by Entrez gene IDs.

Identification of Relapse-Associated PPI Modules

The sEV proteins from PEs were mapped to the reference PPI network. The largest component of the mapped network was used for subsequent analysis, including 568 proteins and 1,632 interactions. This network was integrated with the gene expression profile of the training dataset (GEO: GSE13213). PPI modules (sub-networks) that could discriminate recurrent status were used as previously described.[29] In a given module M with m genes, the expression score (e) of M in sample j was defined as:

$$e_j = \sum_{i \in m} \frac{Z_{ij}}{\sqrt{m}},$$

where Z_{ij} is the Z-transformed gene expression value of gene i.

Then, the discriminative score S(M) of module M was defined as the mutual information (MI) between e and relapse class (c):

$$S(M) = MI(e', c) = \sum_{x \in e'} \sum_{y \in c} p(x, y) \log \frac{p(x, y)}{p(x)p(y)},$$

where e' represents the discretized form of e, by discretizing expression score into 8 $(\log_2(n) + 1)$ (where n is sample numbers) equally spaced bins as described previously.[47]

A greedy searching program was performed to identify modules with locally maximal recurrent score S(M) using the similar parameters as in the previous study.[29] Briefly, each protein in the PPI

Table 1. Clinical Characteristics in Five Datasets of Lung ADC

	Training	Validation			
Variables	GEO: GSE13213	GEO: GSE50081	GEO: GSE30219	GEO: GSE31210	GEO: GSE14814
Age					
< 60 ys	43 (37.07)	19 (14.96)	46 (55.42)	96 (42.48)	35 (49.3)
≥ 60 ys	73 (62.93)	108 (85.04)	37 (44.58)	130 (57.52)	36 (50.7)
Gender					
Female	57 (49.14)	62 (48.82)	18 (21.69)	121 (53.54)	34 (47.89)
Male	59 (50.86)	65 (51.18)	65 (78.31)	105 (46.46)	37 (52.11)
Smoking					
Never	56 (48.28)	23 (18.11)	–	115 (50.88)	–
Ever	60 (51.72)	92 (72.44)	–	111 (49.12)	–
N/A	0 (0)	12 (9.45)	–	0 (0)	–
Stage					
I	78 (67.24)	92 (72.44)	79 (95.18)	168 (74.34)	42 (59.15)
II	13 (11.21)	35 (27.56)	3 (3.61)	58 (25.66)	29 (40.85)
III	25 (21.55)	0 (0)	1 (1.20)	0 (0)	0 (0)
EGFR					
Mutation	45 (38.79)	–	–	127 (56.19)	–
Wild type	71 (61.21)	–	–	99 (43.81)	–
KRAS					
Mutation	14 (12.07)	–	–	–	–
Wild type	102 (87.93)	–	–	–	–
TP53					
Mutation	38 (32.76)	–	–	–	–
Wild type	77 (66.38)	–	–	–	–
N/A	1 (0.86)	–	–	–	–
Postsurgery Recurrence					
No	58 (50)	87 (68.50)	56 (67.47)	162 (71.68)	–
Yes	58 (50)	37 (29.13)	27 (32.53)	64 (28.32)	–
N/A	0 (0)	3 (2.36)	0 (0)	0 (0)	–
ACT					
No	–	–	–	–	32 (45.07)
Yes	–	–	–	–	39 (54.93)

The table shows the number (%) of patients in different groups. ys, years; ACT, adjuvant chemotherapy; N/A, not available.

network (568 proteins) was seeded and iteratively expanded with a specified distance d (d = 2 in this study). For each expansion, the protein that yielded the largest improvement of S(M) among all candidate proteins was added to the module. The module expansion stopped when further protein addition did not increase the S(M) by a specified improvement rate of 10%. Modules with only one protein were removed, and modules with >80% common proteins were merged.

Three permutation methods were utilized to estimate the significance of a candidate module M (with m genes): (1) stimulating random PPI modules with m proteins in the network; (2) shuffling the expression value of m genes; (3) shuffling the recurrent class of samples. Each method was performed by 10,000 times. The random discriminative scores were used as null distribution of S(M). PPI modules with a p value < 0.001 by all three methods were considered to be significant. The gene expression score of candidate modules was used in the following modeling process.

Development of Predictive Signature for Recurrence

To construct a PPI network-based gene signature, the gene expression score (as defined above) of candidate modules was used. Feature selection and modeling was performed based on RF algorithm using the R package "varSelRF".[48] In detail, the predictive importance for

recurrence of each candidate module was estimate by an initial RF with 5,000 trees. A stepwise backward selection procedure was used to identify the optimal combination of candidate modules for recurrence prediction. In each iteration, 10% of features were excluded, and the remaining features were used to build an RF with 3,000 trees. This program was stopped when there were only two features left. The final RF model was selected with the smallest feature number that achieved the minimum OOB error among all iterative results. The recurrence risk of patients was defined by the OOB probability in the final RF model.

To assess the efficiency of the PPI network-based methodology, we established a gene-based RF prediction model using the same program and parameters as used in the sEV network-based method. The predictive results of the two methods were compared by ROC curve analysis. The AUROCs with 95% confidence interval (CI) were calculated by the R package "pROC".[49] Also, the two methods were compared by the Harrell's C-index in predicting RFS and OS in the independent datasets. A higher C-index means more accuracy in survival predicting.

Estimation of Tumor Microenvironment Status

A hypoxia metagene across different cancer types was obtained from a previous study.[32] A core angiogenesis signature for primary tumor was identified by Masiero et al.[50] A set of inflammatory cytokines[51] was used to estimate the intratumoral inflammation level. The intratumoral hypoxia, angiogenesis, and inflammation scores were calculated by averaging the *Z*-normalized expression value of the corresponding signature genes as we described previously.[52] The abundance of immune and non-immune cells in the tumor microenvironment was calculated by MCP-counter, a robust estimator for tissue-infiltrating cell populations by gene expression profiles.[27]

Other Statistical Methods

GO and KEGG pathway enrichment analysis for PPI modules was performed by the online tool DAVID.[53] The significance of overlapping between two gene sets was calculated by hypergeometric test. The association between two category variables of patients was calculated by Chi-squared test or Fisher's exact test. GSEA of was used to compare interested gene sets against the patient subgroups classified by 3-PPI-Mod.[54] The correlations of PPI modules with hypoxia, angiogenesis, and inflammation scores as well as the stromal cell abundance were calculated by Pearson's correlation. Multiple testing was adjusted by false discovery rate (FDR) using the Benjamini and Hochberg's method. A comparison of patient survival between low- and high-risk groups assigned by 3-PPI-Mod was performed by Kaplan-Meier curve and log-rank test. DSS and OS of patients in the ACT arm and the OBS were compared by Kaplan-Meier curve and log-rank test, with stratification by the 3-PPI-Mod signature. Multivariate Cox's models were performed to assess the prognostic and predictive value of 3-PPI-Mod signature. All statistical analyses were performed by R software (version 3.3.1). A p value < 0.05 was considered to be significant.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2019.07.021.

AUTHOR CONTRIBUTIONS

Conception and design: J.L. and Q.L.; Development of methodology: B.C. and W.D.; Acquisition of data: S.M., Q.W., M.L., H.L., and T.C.; Analysis and interpretation of data: B.C., W.D., and S.M.; Writing, review, and/or revision of the manuscript: B.C., J.L., Q.L., G.Z., and X.Y.; Administrative, technical, or material support: G.Z. and X.Y.; Study supervision: Q.W., M.L., and H.L. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

The authors are grateful for all the subjects who participated in the study. This work was supported by the National Natural Science Foundation of China (grant numbers 81602731, 81672690, and 81772900), the National Key Research and Development Program of China (grant number 2017YFC0113101), and the Key Research Project of Science and Technology Department of Sichuan Province (grant numbers 2015SZ0053 and 2017SZ0013).

REFERENCES

1. Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. CA Cancer J. Clin. *65*, 87–108.
2. Molina, J.R., Yang, P., Cassivi, S.D., Schild, S.E., and Adjei, A.A. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin. Proc. *83*, 584–594.
3. Consonni, D., Pierobon, M., Gail, M.H., Rubagotti, M., Rotunno, M., Goldstein, A., Goldin, L., Lubin, J., Wacholder, S., Caporaso, N.E., et al. (2015). Lung cancer prognosis before and after recurrence in a population-based setting. J. Natl. Cancer Inst. *107*, djv059.
4. Akagi, I., Okayama, H., Schetter, A.J., Robles, A.I., Kohno, T., Bowman, E.D., Kazandjian, D., Welsh, J.A., Oue, N., Saito, M., et al. (2013). Combination of protein coding and noncoding gene expression as a robust prognostic classifier in stage I lung adenocarcinoma. Cancer Res. *73*, 3821–3832.
5. Zhu, C.Q., Ding, K., Strumpf, D., Weir, B.A., Meyerson, M., Pennell, N., Thomas, R.K., Naoki, K., Ladd-Acosta, C., Liu, N., et al. (2010). Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. J. Clin. Oncol. *28*, 4417–4424.
6. Tkach, M., and Théry, C. (2016). Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. Cell *164*, 1226–1232.
7. Mathieu, M., Martin-Jaular, L., Lavieu, G., and Théry, C. (2019). Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. Nat. Cell Biol. *21*, 9–17.
8. Théry, C., Witwer, K.W., Aikawa, E., Alcaraz, M.J., Anderson, J.D., Andriantsitohaina, R., Antoniou, A., Arab, T., Archer, F., Atkin-Smith, G.K., et al. (2018). Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. J. Extracell. Vesicles *7*, 1535750.
9. Becker, A., Thakur, B.K., Weiss, J.M., Kim, H.S., Peinado, H., and Lyden, D. (2016). Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. Cancer Cell *30*, 836–848.
10. Kalluri, R. (2016). The biology and function of exosomes in cancer. J. Clin. Invest. *126*, 1208–1215.

11. Muller, L., Simms, P., Hong, C.S., Nishimura, M.I., Jackson, E.K., Watkins, S.C., and Whiteside, T.L. (2017). Human tumor-derived exosomes (TEX) regulate Treg functions via cell surface signaling rather than uptake mechanisms. OncoImmunology *6*, e1261243.

12. Boelens, M.C., Wu, T.J., Nabet, B.Y., Xu, B., Qiu, Y., Yoon, T., Azzam, D.J., Twyman-Saint Victor, C., Wiemann, B.Z., Ishwaran, H., et al. (2014). Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. Cell *159*, 499–513.

13. Li, L., Li, C., Wang, S., Wang, Z., Jiang, J., Wang, W., Li, X., Chen, J., Liu, K., Li, C., and Zhu, G. (2016). Exosomes Derived from Hypoxic Oral Squamous Cell Carcinoma Cells Deliver miR-21 to Normoxic Cells to Elicit a Prometastatic Phenotype. Cancer Res. *76*, 1770–1780.

14. Dorayappan, K.D.P., Wanner, R., Wallbillich, J.J., Saini, U., Zingarelli, R., Suarez, A.A., Cohn, D.E., and Selvendiran, K. (2018). Hypoxia-induced exosomes contribute to a more aggressive and chemoresistant ovarian cancer phenotype: a novel mechanism linking STAT3/Rab proteins. Oncogene *37*, 3806–3821.

15. Wang, T., Gilkes, D.M., Takano, N., Xiang, L., Luo, W., Bishop, C.J., Chaturvedi, P., Green, J.J., and Semenza, G.L. (2014). Hypoxia-inducible factors and RAB22A mediate formation of microvesicles that stimulate breast cancer invasion and metastasis. Proc. Natl. Acad. Sci. USA *111*, E3234–E3242.

16. Hsu, Y.L., Hung, J.Y., Chang, W.A., Lin, Y.S., Pan, Y.C., Tsai, P.H., Wu, C.Y., and Kuo, P.L. (2017). Hypoxic lung cancer-secreted exosomal miR-23a increased angiogenesis and vascular permeability by targeting prolyl hydroxylase and tight junction protein ZO-1. Oncogene *36*, 4929–4942.

17. Liu, Y., Gu, Y., Han, Y., Zhang, Q., Jiang, Z., Zhang, X., Huang, B., Xu, X., Zheng, J., and Cao, X. (2016). Tumor Exosomal RNAs Promote Lung Pre-metastatic Niche Formation by Activating Alveolar Epithelial TLR3 to Recruit Neutrophils. Cancer Cell *30*, 243–256.

18. Jin, X., Chen, Y., Chen, H., Fei, S., Chen, D., Cai, X., Liu, L., Lin, B., Su, H., Zhao, L., et al. (2017). Evaluation of Tumor-Derived Exosomal miRNA as Potential Diagnostic Biomarkers for Early-Stage Non-Small Cell Lung Cancer Using Next-Generation Sequencing. Clin. Cancer Res. *23*, 5311–5319.

19. Kanaoka, R., Iinuma, H., Dejima, H., Sakai, T., Uehara, H., Matsutani, N., and Kawamura, M. (2018). Usefulness of Plasma Exosomal MicroRNA-451a as a Noninvasive Biomarker for Early Prediction of Recurrence and Prognosis of Non-Small Cell Lung Cancer. Oncology *94*, 311–323.

20. Sandfeld-Paulsen, B., Jakobsen, K.R., Bæk, R., Folkersen, B.H., Rasmussen, T.R., Meldgaard, P., Varming, K., Jørgensen, M.M., and Sorensen, B.S. (2016). Exosomal Proteins as Diagnostic Biomarkers in Lung Cancer. J. Thorac. Oncol. *11*, 1701–1710.

21. Sandfeld-Paulsen, B., Aggerholm-Pedersen, N., Bæk, R., Jakobsen, K.R., Meldgaard, P., Folkersen, B.H., Rasmussen, T.R., Varming, K., Jørgensen, M.M., and Sorensen, B.S. (2016). Exosomal proteins as prognostic biomarkers in non-small cell lung cancer. Mol. Oncol. *10*, 1595–1602.

22. Ueda, K., Ishikawa, N., Tatsuguchi, A., Saichi, N., Fujii, R., and Nakagawa, H. (2014). Antibody-coupled monolithic silica microtips for highthroughput molecular profiling of circulating exosomes. Sci. Rep. *4*, 6232.

23. Park, J.O., Choi, D.Y., Choi, D.S., Kim, H.J., Kang, J.W., Jung, J.H., Lee, J.H., Kim, J., Freeman, M.R., Lee, K.Y., et al. (2013). Identification and characterization of proteins isolated from microvesicles derived from human lung cancer pleural effusions. Proteomics *13*, 2125–2134.

24. Wu, A.Y., Ueda, K., and Lai, C.P. (2019). Proteomic Analysis of Extracellular Vesicles for Cancer Diagnostics. Proteomics *19*, e1800162.

25. Li, Y., Zhang, Y., Qiu, F., and Qiu, Z. (2011). Proteomic identification of exosomal LRG1: a potential urinary biomarker for detecting NSCLC. Electrophoresis *32*, 1976–1983.

26. Clark, D.J., Fondrie, W.E., Yang, A., and Mao, L. (2016). Triple SILAC quantitative proteomic analysis reveals differential abundance of cell signaling proteins between normal and lung cancer-derived exosomes. J. Proteomics *133*, 161–169.

27. Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. *17*, 218.

28. Costa-Silva, B., Aiello, N.M., Ocean, A.J., Singh, S., Zhang, H., Thakur, B.K., Becker, A., Hoshino, A., Mark, M.T., Molina, H., et al. (2015). Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat. Cell Biol. *17*, 816–826.

29. Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D., and Ideker, T. (2007). Network-based classification of breast cancer metastasis. Mol. Syst. Biol. *3*, 140.

30. Cao, B., Luo, L., Feng, L., Ma, S., Chen, T., Ren, Y., Zha, X., Cheng, S., Zhang, K., and Chen, C. (2017). A network-based predictive gene-expression signature for adjuvant chemotherapy benefit in stage II colorectal cancer. BMC Cancer *17*, 844.

31. Stuelten, C.H., Parent, C.A., and Montell, D.J. (2018). Cell motility in cancer invasion and metastasis: insights from simple model organisms. Nat. Rev. Cancer *18*, 296–312.

32. Buffa, F.M., Harris, A.L., West, C.M., and Miller, C.J. (2010). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. Br. J. Cancer *102*, 428–435.

33. Müller, C.B., de Barros, R.L., Castro, M.A., Lopes, F.M., Meurer, R.T., Roehe, A., Mazzini, G., Ulbrich-Kulczynski, J.M., Dal-Pizzol, F., Fernandes, M.C., et al. (2011). Validation of cofilin-1 as a biomarker in non-small cell lung cancer: application of quantitative method in a retrospective cohort. J. Cancer Res. Clin. Oncol. *137*, 1309–1316.

34. Castro, M.A., Dal-Pizzol, F., Zdanov, S., Soares, M., Müller, C.B., Lopes, F.M., Zanotto-Filho, A., da Cruz Fernandes, M., Moreira, J.C., Shacter, E., and Klamt, F. (2010). CFL1 expression levels as a prognostic and drug resistance marker in non-small cell lung cancer. Cancer *116*, 3645–3655.

35. Delorme-Walker, V., Seo, J.Y., Gohla, A., Fowler, B., Bohl, B., and DerMardirossian, C. (2015). Chronophin coordinates cell leading edge dynamics by controlling active cofilin levels. Proc. Natl. Acad. Sci. USA *112*, E5150–E5159.

36. Teng, Y., Qin, H., Bahassan, A., Bendzunas, N.G., Kennedy, E.J., and Cowell, J.K. (2016). The WASF3-NCKAP1-CYFIP1 Complex Is Essential for Breast Cancer Metastasis. Cancer Res. *76*, 5133–5142.

37. Guo, C., Liu, S., Wang, J., Sun, M.Z., and Greenaway, F.T. (2013). ACTB in cancer. Clin. Chim. Acta *417*, 39–44.

38. Rasheed, S.A., Teo, C.R., Beillard, E.J., Voorhoeve, P.M., Zhou, W., Ghosh, S., and Casey, P.J. (2015). MicroRNA-31 controls G protein alpha-13 (GNA13) expression and cell invasion in breast cancer cells. Mol. Cancer *14*, 67.

39. Zheng, B., Liang, L., Huang, S., Zha, R., Liu, L., Jia, D., Tian, Q., Wang, Q., Wang, C., Long, Z., et al. (2012). MicroRNA-409 suppresses tumour cell invasion and metastasis by directly targeting radixin in gastric cancers. Oncogene *31*, 4509–4516.

40. Wang, P., Deng, Y., and Fu, X. (2017). MiR-509-5p suppresses the proliferation, migration, and invasion of non-small cell lung cancer by targeting YWHAG. Biochem. Biophys. Res. Commun. *482*, 935–941.

41. Rhee, I., and Veillette, A. (2012). Protein tyrosine phosphatases in lymphocyte activation and autoimmunity. Nat. Immunol. *13*, 439–447.

42. Salmond, R.J., Filby, A., Qureshi, I., Caserta, S., and Zamoyska, R. (2009). T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance. Immunol. Rev. *228*, 9–22.

43. Kumar, V., Cheng, P., Condamine, T., Mony, S., Languino, L.R., McCaffrey, J.C., Hockstein, N., Guarino, M., Masters, G., Penman, E., et al. (2016). CD45 Phosphatase Inhibits STAT3 Transcription Factor Activity in Myeloid Cells and Promotes Tumor-Associated Macrophage Differentiation. Immunity *44*, 303–315.

44. True, L.D., Zhang, H., Ye, M., Huang, C.Y., Nelson, P.S., von Haller, P.D., Tjoelker, L.W., Kim, J.S., Qian, W.J., Smith, R.D., et al. (2010). CD90/THY1 is overexpressed in prostate cancer-associated fibroblasts and could serve as a cancer biomarker. Mod. Pathol. *23*, 1346–1356.

45. Kowal, J., Arras, G., Colombo, M., Jouve, M., Morath, J.P., Primdal-Bengtson, B., Dingli, F., Loew, D., Tkach, M., and Théry, C. (2016). Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc. Natl. Acad. Sci. USA *113*, E968–E977.

46. Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T.K., Chandrika, K.N., Deshpande, N., Suresh, S., et al. (2004). Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. *32*, D497–D501.

47. Tourassi, G.D., Frederick, E.D., Markey, M.K., and Floyd, C.E., Jr. (2001). Application of the mutual information criterion for feature selection in computer-aided diagnosis. Med. Phys. *28*, 2394–2402.

48. Díaz-Uriarte, R., and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. BMC Bioinformatics *7*, 3.

49. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 77.

50. Masiero, M., Simões, F.C., Han, H.D., Snell, C., Peterkin, T., Bridges, E., Mangala, L.S., Wu, S.Y., Pradeep, S., Li, D., et al. (2013). A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis. Cancer Cell *24*, 229–241.

51. Saloura, V., Zuo, Z., Koeppen, H., Keck, M.K., Khattri, A., Boe, M., Hegde, P.S., Xiao, Y., Nakamura, Y., Vokes, E.E., et al. (2014). Correlation of T-cell inflamed phenotype with mesenchymal subtype, expression of PD-L1, and other immune checkpoints in head and neck cancer. J. Clin. Oncol. *32* (*Suppl 15*), 6009–6009.

52. Cao, B., Wang, Q., Zhang, H., Zhu, G., and Lang, J. (2017). Two immune-enhanced molecular subtypes differ in inflammation, checkpoint signaling and outcome of advanced head and neck squamous cell carcinoma. OncoImmunology *7*, e1392427.

53. Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

54. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.