



Integrative Analysis of Membrane Proteome and MicroRNA Reveals Novel Lung Cancer Metastasis Biomarkers

Yan Kong^{1†}, Zhi Qiao^{2†}, Yongyong Ren¹, Georgi Z. Genchev^{1,3,4}, Maolin Ge⁵, Hua Xiao^{2*}, Hongyu Zhao^{6*} and Hui Lu^{1,2,3*}

OPEN ACCESS

Edited by:

Jun Ding,
Carnegie Mellon University,
United States

Reviewed by:

Khanh N. Q. Le,
Taipei Medical University, Taiwan
Yuqing Liu,
Hematogenix Laboratory Services,
LLC, United States

*Correspondence:

Hua Xiao
huaxiao@sjtu.edu.cn
Hongyu Zhao
hongyu.zhao@yale.edu
Hui Lu
huilu@sjtu.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 18 May 2020

Accepted: 11 August 2020

Published: 28 August 2020

Citation:

Kong Y, Qiao Z, Ren Y,
Genchev GZ, Ge M, Xiao H, Zhao H
and Lu H (2020) Integrative Analysis
of Membrane Proteome
and MicroRNA Reveals Novel Lung
Cancer Metastasis Biomarkers.
Front. Genet. 11:1023.
doi: 10.3389/fgene.2020.01023

¹ SJTU-Yale Joint Center for Biostatistics and Data Science, Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ² State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ³ Center for Biomedical Informatics, Shanghai Engineering Research Center for Big Data in Pediatric Precision Medicine, Shanghai Children's Hospital, Shanghai, China, ⁴ Bulgarian Institute for Genomics and Precision Medicine, Sofia, Bulgaria, ⁵ State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital, School of Medicine and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ⁶ Department of Biostatistics, Yale University, New Haven, CT, United States

Lung cancer is one of the most common human cancers both in incidence and mortality, with prognosis particularly poor in metastatic cases. Metastasis in lung cancer is a multifarious process driven by a complex regulatory landscape involving many mechanisms, genes, and proteins. Membrane proteins play a crucial role in the metastatic journey both inside tumor cells and the extra-cellular matrix and are a viable area of research focus with the potential to uncover biomarkers and drug targets. In this work we performed membrane proteome analysis of highly and poorly metastatic lung cells which integrated genomic, proteomic, and transcriptional data. A total of 1,762 membrane proteins were identified, and within this set, there were 163 proteins with significant changes between the two cell lines. We applied the Tied Diffusion through Interacting Events method to integrate the differentially expressed disease-related microRNAs and functionally dys-regulated membrane protein information to further explore the role of key membrane proteins and microRNAs in multi-omics context. *Has-miR-137* was revealed as a key gene involved in the activity of membrane proteins by targeting MET and PXN, affecting membrane proteins through protein-protein interaction mechanism. Furthermore, we found that the membrane proteins CDH2, EGFR, ITGA3, ITGA5, ITGB1, and CALR may have significant effect on cancer prognosis and outcomes, which were further validated *in vitro*. Our study provides multi-omics-based network method of integrating microRNAs and membrane proteome information, and uncovers a differential molecular signatures of highly and poorly metastatic lung cancer cells; these molecules may serve as potential targets for giant-cell lung metastasis treatment and prognosis.

Keywords: membrane proteome, lung cancer metastasis, multi-omics analysis, microRNA, prognostic

INTRODUCTION

Lung cancer remains the leading cause of cancer-related mortality in the world, with an overall 5-year survival rate of 18% (Siegel et al., 2017). In 2016, over 155,000 people died from lung cancer in the United States alone (Siegel et al., 2017). These low survival rates are partly due to the fact that over 50% of patients are diagnosed at a later stage, for which the 5-year survival is only 4%. Approximately 80–85% of lung cancer cases are non-small cell lung cancer (NSCLC), and the remaining 15–20% are small-cell lung cancer cases (Giaccone and Zucali, 2008). In NSCLC, it is estimated that over 40% of patients have metastases at the time of diagnosis (Waqar et al., 2018). The prognosis is poor in metastatic cases -only ~1% of such NSCLC patients will survive five or more years (Borghaei et al., 2017).

Cancer metastasis involves tumor cell invasion across interstitial tissues and basement membranes (Jiang et al., 2015; Qiao et al., 2019). Abnormal expression of membrane proteins in cancer tissues and cells has been shown to play a key role in cancer occurrence and metastasis (Leth Larsen et al., 2009; Kampen, 2011). In a recent study of hepatocellular carcinoma (HCC), golgi membrane protein 1 (GOLM1) was shown as a key target of miR-382, HCC cells metastasis status was inhibited when GOLM1 is down-regulated in HCC cells (Zhang et al., 2018). *N*-cadherin (CDH2), a direct target of miR-145, is a cell-cell adhesion molecule that contributes to the invasive/metastatic phenotype in many cancers such as gastric cancer, breast cancer, and lung cancer (Lei et al., 2017; Mo et al., 2017; Ye et al., 2018). β -catenin (CTNNB1), involved in the regulation of cell adhesion, promote ovarian cancer metastasis and liver cancer (Arend et al., 2013; Ding et al., 2014). Mucin 1 (MUC1) protein, a membrane-tethered mucin glycoprotein, is also associated with poor prognosis and enhanced metastasis in human pancreatic cancers (Wu et al., 2018). In addition, microRNAs have important roles in cancer metastasis (Nicoloso et al., 2009), and multiple microRNAs, such as hsa-miR-1, hsa-miR-217, hsa-miR-206, and has-miR-577 were previously shown to play key roles in cancer metastasis (Liu et al., 2015; Chen et al., 2017; Samaeekia et al., 2017).

The complex regulatory landscape of cancer metastasis underscores the need of integrative approaches in cancer research. Multi-omics computational studies are an active area of investigation and perform analysis of genomic, proteomic, and transcriptional data combined with prior knowledge of regulatory relationships to uncover clinically relevant discovery such as biomarkers, therapeutically targets, and outcome prediction (Liu et al., 2017; Jiang et al., 2019; Xu et al., 2020). One such recently proposed method – the Tied Diffusion Through Interacting Events (TieDIE) algorithm uses a network diffusion approach to connect genomic perturbations to transcriptional changes (Paull et al., 2013). With the help of the TieDIE algorithm, a contributing factor (small GTPase RHEB) to the differences observed between BRAF and RAS mutants was discovered; in another cancer study, researchers combined transcriptional regulators, mutated genes, and differentially expressed kinases with TieDIE and synthesized a robust

signaling network which consists of drug-able kinase pathways (Drake et al., 2016).

Although both membrane proteome and microRNA have been shown of great importance previously, there was no systematic combined analysis using membrane proteome together with microRNA data on lung squamous cell carcinoma (LUSC). In this study, we utilized quantitative membrane proteome and microRNA expression together with multiple regulation networks to perform comparative analysis of highly and poorly metastatic lung cancer cell lines (95C and 95D). In the following, we described the methodology used for experiment and computational analysis; differentially expressed membrane proteins are identified, then using joint analysis method to integrate microRNA expression data. Finally the significance of the study is discussed.

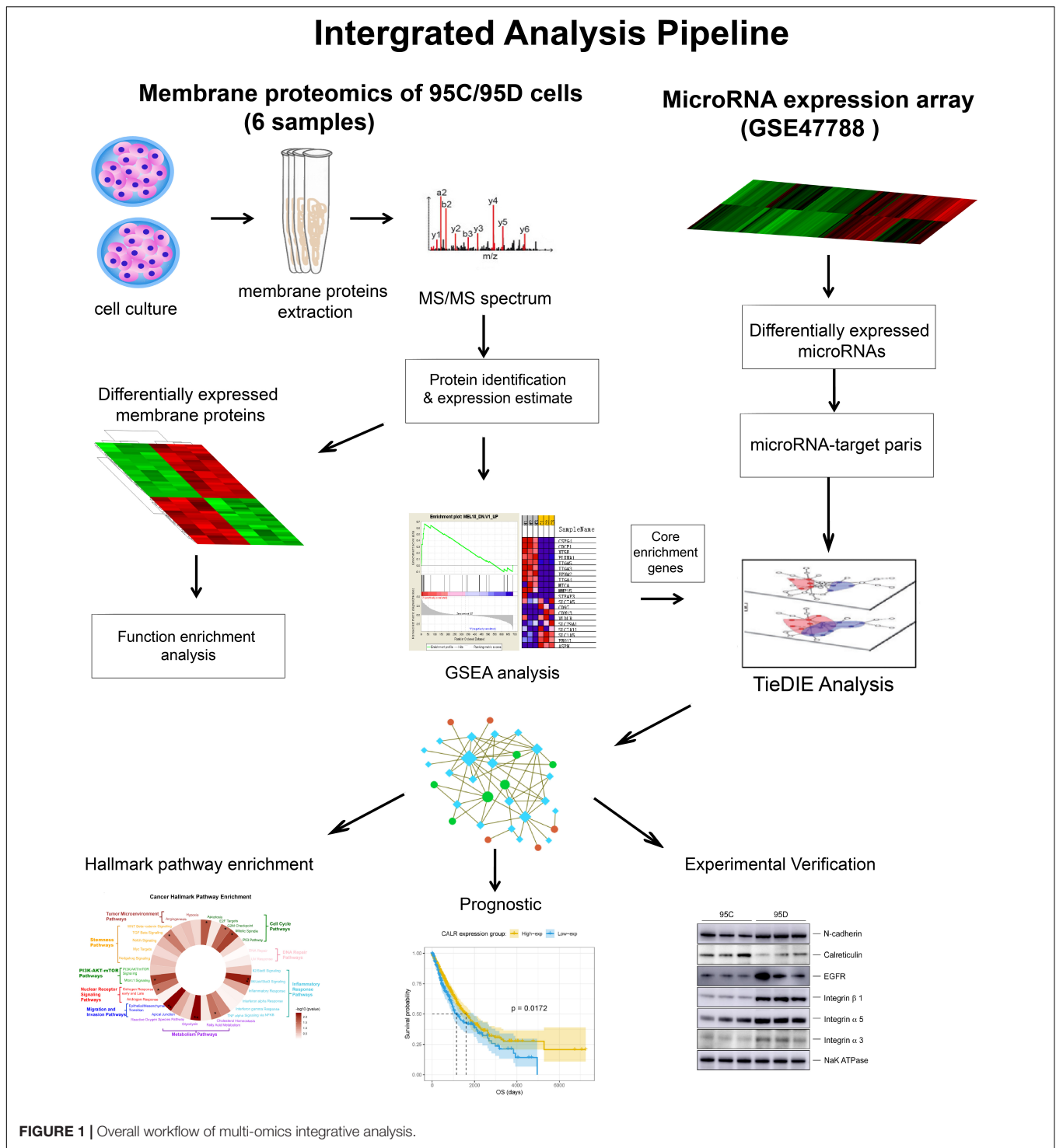
MATERIALS AND METHODS

The methods utilized in the work aim to discover biomarkers which are associated with disease outcomes measured by overall survival (OS). We integrated experimental and computational approaches, proteomics and genomics (microRNA) data, and bioinformatics analysis to drive bio-medical discovery. The overall work-flow is depicted in **Figure 1**.

Experimental Data Collection – Protein Identification and Quantification

Cell Culture and Membrane Protein Preparation

Human giant cell lung cancer cell lines of poorly (95C) and highly (95D) metastatic potential were purchased from the Institute of Biochemistry and Cell Biology of the Chinese Academy of Sciences (Shanghai, China). Cells were cultured in DMEM and RPMI 1640 medium supplemented with 10% fetal bovine serum for 95D and 95C cells respectively at 37°C and 5% CO₂ incubator. All culture media were supplemented with 100 U/mL penicillin and 100 mg/mL streptomycin sulfate. All cultured cells were tested for mycoplasma contamination before use. Membrane proteins isolation was performed with the Pierce® Cell Surface Protein Isolation Kit (Pierce, Thermo Fisher Scientific, United States), and followed the protocol described by de Wit et al. (2012). All reagents were cooled to 4°C before protein biotinylation. The cells were washed twice with ice-cold phosphate buffered saline (PBS) followed by incubation with 0.25 mg/mL Sulfo-NHS-SS-biotin (Pierce) in 48 mL ice-cold PBS per flask on orbital shaker for 30 min at 4°C. Then, 500 μ L of quenching solution were added to each flask to quench the reaction. After being washed with ice-cold PBS, harvested by gentle scraping, and pelleted by centrifugation, the cells were lysed using the Lysis Buffer (Pierce) which was added with protease inhibitors for 30 min on ice with vortexing every 5 min for 5 s. The cell lysates were centrifuged at 10,000 \times g for 2 min at 4°C to remove cell remnants. Before clarified supernatant was used to purify biotinylated proteins on NeutrAvidin Agarose (Pierce), 500 μ L of NeutrAvidin Agarose slurry was added and centrifuged 1 min at 1,000 \times g and the flow-through was discarded followed by



washing with Pierce Wash Buffer in a provided column (Pierce) thrice. The clarified supernatant was added and incubated for 2 h at 4°C using an end-over-end tumbler to mix vigorously and allow the biotinylated proteins to bind to the NeutrAvidin Agarose slurry. Unbound proteins were removed by washing with 1% Non-idet-P40 and 0.1% SDS in 500 μL PBS thrice and then by washing with 0.1% Non-idet-P40 and 0.5 M

NaCl in 500 μL PBS thrice. Finally, the captured proteins were eluted from the biotin-NeutrAvidin Agarose and were collected by column centrifugation at 1,000 × g for 2 min. Three biological replicates were obtained for both cell lines. All protein concentrations were quantified using the BCA protein Assay Kit (Pierce) and the lysates were stored at -20°C for further analysis.

Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS)

Equal amount of proteins was digested overnight at 37°C by trypsin (Promega, Madison, WI, United States) using the FASP approach. Briefly, 30 µg membrane proteins were used and three biological repeats of each cell line were prepared. Equal amount of peptide was injected into Easy-nLC 1,000 m (Thermo Scientific) coupled with a Q-Exactive mass spectrometer (Thermo Fisher Scientific) (Michalski et al., 2011; Kelstrup et al., 2012). Peptides were eluted to analytical column (75 µm × 15 cm) packed with Jupiter Proteo resin (3 µm, C18, 300 Å, Phenomenex, Torrance, CA, United States). The mobile phase consisted of buffer A (2% acetonitrile and 0.1% formic acid in water) and buffer B (0.1% formic acid in 95% acetonitrile). A flow rate of 250 nL/min and 60 min of the gradient from 12% B to 32% B was applied for the separation of peptides. MS Scan range was from 300 to 1,600 m/z with the resolution of 70,000. For MS/MS, scan range was from 200 to 2,000 m/z with the resolution of 17,500. We performed full MS scan in a positive mode and then selected the five most dominant ions from the initial MS scan for collision-induced dissociation.

Protein Identification and Quantitation

To identify proteins from the acquired data, MS/MS spectra were searched against the Human SwissProt database (548, 208 sequences) (Jungo et al., 2012) using the MASCOT software (version 2.0) (Matrix Science, London, United Kingdom). SwissProt database is a high quality manually annotated and non-redundant protein sequence database, and now more and more bioinformatics data mining algorithms are designed using this database (Le et al., 2017a, 2019a).

The parameters for searching were the MASCOT defaults – enzyme of trypsin, two missed cleavage, fixed modifications of carbamidomethyl (C), and variable modifications of oxidation (M). We set mass tolerance to 20 ppm for MS precursors and 0.05 Da for fragment ions, and then peptide charges of +2, +3, and +4 were retained. For protein identification, we used *p*-value less than 0.05 and false discovery rate (FDR) less than 0.01 at the protein level as the criteria to distinguish two peptides. Label-free quantification was performed by intensity-based absolute quantification (iBAQ), which was based on at least two unique peptides to quantify the different protein profiling in the 95C cell and 95D cell membrane. Quantile normalization was performed to ensure that each sample had the same distribution, the two-fold change and *p*-value less than 0.01 cut-off was set up for the screening of differentially expressed proteins.

Bioinformatics Analysis

Protein Subcellular Localization Annotation and Transmembrane Domain Prediction

The first step of the bioinformatics analysis aimed to identify membrane proteins. Here, we used the following process to annotate the membrane proteins. First, Gene Ontology (GO) cellular component annotation of all identified proteins was performed by the R *go.db* package. The GO is a human and machine readable gene annotation resource, which has been widely used to enable computational discovery in diverse

areas such as protein function identification (Le et al., 2017b, 2019b), text mining in life sciences (Przybyla et al., 2016), and underlying molecular disease mechanisms (Kramarz et al., 2020). Second, additional subcellular location information was downloaded from the UniProt database (Jungo et al., 2012) and added to the protein subcellular localization annotation. Third, transmembrane domains in all identified membrane proteins were predicted by TMHMM¹ Serve v.2.0 (Krogh et al., 2001). This gave us a list of 3,240 membrane proteins which were utilized in the downstream analysis.

Differential Expression Analysis and Enrichment Analysis

In the second step of the bioinformatics analysis we searched for protein differential expression and enrichment. Utilizing the Student's *t*-test, we tested the 3,240 membrane proteins for differential expression between poorly metastatic 95C and highly metastatic 95D cell lines. Proteins with a *p*-value <0.05 and $|\log_2FC| > 1$ were considered to be significantly differentially expressed and included in the reduced set (*n* = 163) for further analysis. GO and KEGG pathway enrichment analysis by the *clusterprofiler* R package (Yu et al., 2012) was performed on the differentially expressed membrane proteins in order to understand which function they may affect. Gene set enrichment analysis (GSEA) were performed by the GSEA software v.3.0 (Subramanian et al., 2005) using the molecular signatures database MSigDB (Liberzon et al., 2011). At the end of this step we obtained a list of 163 differentially expressed proteins and 87 enriched genes.

Multi-Omics Data – MicroRNA

To integrate genomics data into our analysis, microRNA expression array data of 95D and 95C cells (GSE47788) was downloaded from the Gene Expression Omnibus (GEO) database (Edgar et al., 2002; Wang X. M. et al., 2013). We directly used the differential expression results provided by the study. For the 64 differentially expressed microRNAs, we applied the miRWALK2.0 (Dweep and Gretz, 2015) software to build a list of microRNAs and their gene targets.

Generation of Biological Pathway Subnetwork Connecting MicroRNA and Enriched Genes

Having obtained (1) the set of 87 enriched genes and (2) the set 64 of differentially expressed microRNA and their gene targets, we build a sub-network which significantly close-connects these genes and microRNAs. We used the *Tied Diffusion through Interacting Events* (TieDIE) software (Paull et al., 2013; Drake et al., 2016). The Multinet pathway database (Brown et al., 2013) together with the validated microRNA-target pairs selected by miRWALK2.0 in the previous step served as the background network of the TieDIE program. 32 membrane protein genes and 180 linker genes were selected in this step.

Cancer Hallmarks Enrichment Calculation

The set of genes in the sub-network built with TieDIE in the previous step were used to perform cancer hallmark enrichment

¹<http://www.cbs.dtu.dk/services/TMHMM/>

which is now a popular functional analysis method for a gene clustering (Drake et al., 2016; Ge et al., 2020). Cancer hallmark definitions were also downloaded from the MSigDB database, and the cancer hallmark pathway enrichment were performed by calculating the probability of overlap between input genes and the hallmark pathway gene sets and evaluated by hyper geometric test with Benjamini and Hochberg (1997) correction of *p*-values. The source code for cancer hallmark enrichment analysis is publicly available at: <https://github.com/YankongSJTU/CHEA>.

Survival Analysis

Integrating the above-obtained gene lists with survival data, we performed patient survival analysis to determine if the selected genes have impact on cancer related outcomes. We downloaded mRNA expression data (FPKM values) for the 32 membrane protein genes from the Human Protein Atlas² (Uhlen et al., 2010) of lung adenocarcinoma (LAC) and LUSC patients (Tomczak et al., 2015), 925 samples in total. From the same source we also obtained patient OS data. For each gene, FPKM values from the 20th to 80th percentiles were used to group the patients; significant differences in the survival outcomes of the groups were examined and the value yielding the lowest log-rank *p*-value was set to be the best cut-off value. Patients were classified into two groups: group 1 with values above the cutoff (high expression level group) and group 2 with values below the cutoff (low expression level group). The outcome differences for each group were calculated using the Log-rank test by the Kaplan–Meier method (Bland and Altman, 1998). Prognostic analysis was performed by using the R packages *KMsurv*, *survival*, and *survminer* (Latouche and Aurelien, 2019). All *p*-values were derived from two-tailed statistical tests, and *p*-value < 0.05 was considered as statistically significant. At the completion of the bioinformatics analysis, six significant gene biomarkers significantly correlated with OS were determined.

We also constructed a prognostic risk index with these selected membrane proteins:

$$\text{Risk Index} = \sum \text{expression}(P_i) * \frac{\text{HR}_i}{\text{maximum}(\text{expression}(P_i))} \quad (1)$$

where P_i ($i = 1, 2, 3, \dots, K$) means the selected K membrane proteins. Samples are grouped according to the risk factor levels, and the prognosis differences are compared.

Experimental Verification

The cellular expression of the six candidate proteins were verified experimentally via Western Blot analysis. Total cell lysates were obtained using RIPA buffer (Thermo Fisher Scientific, United States). Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to polyvinylidene fluoride (PVDF) membrane (Millipore, Burlington, MA, United States). The membranes were blocked in PBS, 10% (w/v) skim milk for 1 h in phosphate buffer saline-Tween 20 (PBS-T), and incubated for 3 h at RT in 5% milk in

PBS-T with primary antibodies: CDH2, EGFR, ITGA3, ITGA5, ITGB1, and CALR (Abcam, Cambridge, United Kingdom). Then, after washing, the PVDF membrane was incubated with secondary antibody (The Jackson Laboratory, United States) for 40 min at RT. Thermo Scientific SuperSignal West Pico PLUS Chemiluminescent Substrate kit (Thermo Fisher Scientific, United States) was used for visualization.

RESULTS AND DISCUSSION

Protein Identification and Differential Expression Analysis

We first sought to determine the full set of membrane proteins detected by mass spectrometry and then identify the subset of differentially expressed ones. A total of 3,241 unique proteins were identified and quantified. A total of 3,107 proteins (95.9%) were annotated by GO cellular component analysis and 2,887 proteins (89.1%) were annotated by using the UniProt subcellular localization database. Subcellular localization annotation analysis predicted that 1,762 proteins were membrane proteins (54.7%), whereas the TMHMM algorithm predicted that 590 proteins (18.3%) had a transmembrane domain. Significant differences between 95D and 95C cells were observed in the expression of 163 membrane proteins ($|\log_2\text{FC}| > 1$ and *p*-value < 0.05). We provide detailed subcellular localization annotations together with TMHMM results of all proteins in **Supplementary Table S1**. All differential expression results are listed in **Supplementary Table S2**.

Functional Characterization of Differentially Expressed Membrane Proteins

Differential expressed membrane proteins may play a key role in tumor metastasis. Membrane proteins as well as extracellular matrix (ECM) molecules, cell adhesion molecules and adhesion receptors form into functional complex units and maintain cell–cell adhesions. These complexes, once disassembled, will increase tumor metastasis and invasion (Gumbiner, 1996; Vadakekolathu et al., 2018). To further examine the mechanistic role in cancer metastasis of differentially expressed membrane proteins, we performed two-step functional enrichment analysis.

First, GO and KEGG functional enrichment analysis revealed that differentially expressed membrane proteins were mainly centered on focal adhesion and cell-substrate adherens junctions, and many metabolic pathways. For example, the three most significantly enriched GO terms were GO:0005925~focal adhesion (*q*-value = $7.2e-43$), GO:0005924~cell-substrate adherens junction (*q*-value = $7.2e-43$), and cell-substrate junction (*q*-value = $1.2e-42$), and the three most significantly enriched KEGG pathway terms were hs03010~Ribosome (*q*-value = $1.07e-16$), hsa05412~Arrhythmogenic right ventricular cardiomyopathy (ARVC) (*q*-value = $1.65e-4$), and hsa05416~Viral myocarditis (*q*-value = $3.22e-3$). All detailed enrichment results are shown in **Supplementary Table S4**.

²<https://www.proteinatlas.org>

In the second step of this analysis, GSEA revealed that differentially expressed membrane proteins were enriched in 13 GO categories (MSigDB c6 Gene Ontology or GO categories) including of GO_POSITIVE_REGULATION_OF_LOCOMOTION and GO_CELL_JUNCTION_ORGANIZATION. KEGG_FOCAL_ADHESION pathway was also enriched. Similar molecular terms, “KEGG_FOCAL_ADHESION” and “REACTOME_HEMOSTASIS” were also identified using other functional gene sets (MSigDB c6 KEGG or Reactome categories) (Table 1). Of all 87 enriched genes, 21 were considered as a core enrichment set, i.e., genes which were considered to contribute the most to the enrichment result According to the differential expression analysis, all core enrichment genes were significantly up-regulated, including ALCAM, LDOA, TP1A1, TP1B1, TP1B3, CDH2, CTNNB1, CXADR, EGFR, EPHA2, GLG1, ITGA3, ITGA5, ITGB1, ITGB3, JAM3, MPZL1, NEGR1, PARK7, PTPRF, and SLC16A3 (see Supplementary Table 5).

MicroRNAs and Significant Sub-Network Derived From TiedIE Program

MicroRNAs are non-coding RNAs which participate in cellular activity by regulating target genes. Increasing number of studies have reported that microRNAs are frequently differentially expressed in numerous types of human cancer and play an important role in the progression and development of NSCLC (Inamura and Ishikawa, 2016; Berrout et al., 2017; Chang et al., 2017; Zhang et al., 2017; Iqbal et al., 2018). Although microRNAs act only in the cell where they are synthesized, they can also influence the functions of neighboring cells or play a role in the tumor micro-environment by modulating the ECM state (Rutnam et al., 2013).

Utilizing the TiedIE program (a pathway-based multi-omics method which extends on the heat diffusion strategies and uses a network diffusion approach to connect proteins and genes related to diseases) we were able to link differentially expressed microRNAs with membrane proteins.

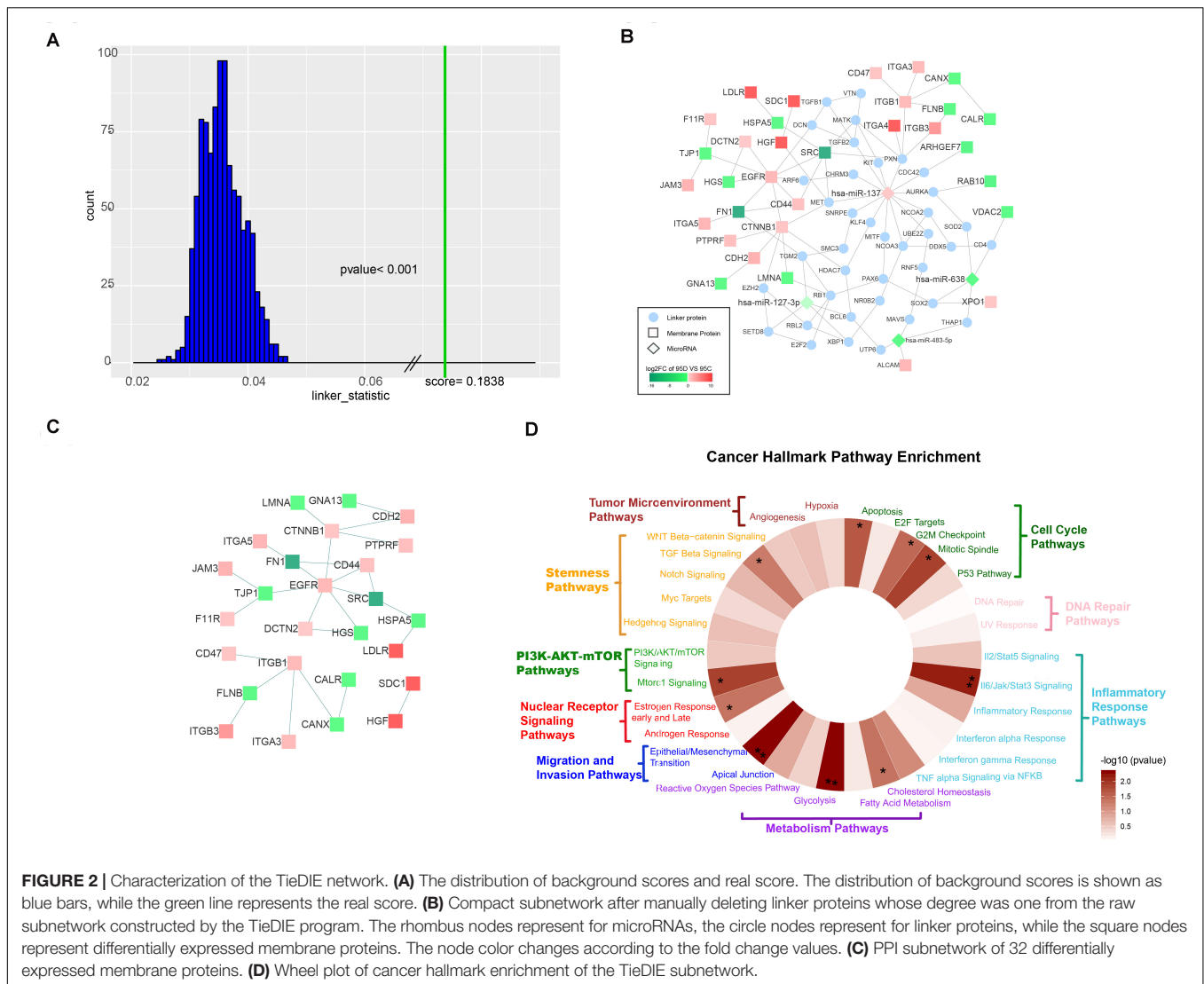
First, 64 differentially expressed microRNAs were obtained directly from a previous study (Supplementary Table S2). We successfully found 300 targets of 64 microRNAs with the help of MIRWALK 2.0 and synthesized a validated microRNA-regulated target network (Supplementary Table S3) which served as background database together with the Multinet database (Khurana et al., 2013). The input membrane proteins and microRNAs were found to be significantly close ($p < 0.001$) in a pathway space based on a background model generated by 1,000 permutations of the data (Paull et al., 2013), where each input set (membrane proteins, microRNAs) was swapped with genes of similar network connectivity while the other one was fixed (Figure 2A).

We selected a compact sub-network with a high level of specificity, which consisted of 216 nodes – four microRNAs (hsa-miR-137, hsa-miR-483-5p, hsa-miR-638, and hsa-miR-127-3p), 32 differentially expressed membrane proteins (HSPA5, CANX, TJP1, FN1, ITGA3, ITGA4, ITGA5, XPO1, JAM3, F11R, SDC1, FLNB, DCTN2, VDAC2, ALCAM, RAB10, SRC, LDLR, CDH2, EGFR, CTNNB1, ITGB1, ITGB3, CD44, CD47, HGS, ARHGEF7, HGF, LMNA, CALR, PTPRF, and GNA13) and 180 linking proteins connected by 244 edges (Supplementary Table S6). We manually deleted those linker proteins whose degree was 1 and constructed a sub-network consisting of four microRNAs, 32 membrane proteins, and 38 linker proteins with 101 edges (Figure 2B). Focusing on differentially expressed membrane proteins, we found 32 membrane proteins which were connected

TABLE 1 | Gene set enrichment result of differentially expressed membrane proteins.

GS(follow, link, to, MSigDB)	Size	ES	NES	NOM, p-value	FDR, q-value
GO_POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	17	0.49060193	2.2359705	0	0.044999983
GO_POSITIVE_REGULATION_OF_LOCOMOTION	15	0.52440554	2.2129607	0	0.044999983
GO_CELL_PROJECTION_PART	18	0.46777532	2.201335	0	0.044999994
GO_PLASMA_MEMBRANE_REGION	24	0.5549008	2.1945791	0	0.044999983
GO_CELLULAR_COMPONENT_MORPHOGENESIS	16	0.51833624	2.1498182	0	0.044999994
KEGG_FOCAL_ADHESION	17	0.45756185	1.9556208	0	0.044999994
REACTOME_HEMOSTASIS	26	0.38783315	1.9427094	0	0.045000017
GO_CELL_JUNCTION_ORGANIZATION	16	0.48022297	1.9331897	0	0.044999983
GO_REGULATION_OF_CELL_DIFFERENTIATION	18	0.49768242	1.9189458	0	0.044999994
GO_MEMBRANE_MICRODOMAIN	15	0.5177578	1.9009451	0	0.044999994
GO_TISSUE_DEVELOPMENT	32	0.38096464	1.8797828	0	0.045
GO_ANATOMICAL_STRUCTURE_FORMATION_INVOLVED_IN_MORPHOGENESIS	18	0.3993144	1.8518116	0	0.044999994
GO_REGULATION_OF_PHOSPHORUS_METABOLIC_PROCESS	23	0.42798108	1.8506685	0	0.045000024
GO_REGULATION_OF_PROTEIN_MODIFICATION_PROCESS	21	0.46240425	1.8467267	0	0.045000017
GO_ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	15	0.5241594	1.8411065	0	0.044999999
GO_RECEPTOR_BINDING	24	0.39534718	1.8257784	0	0.044999983
GO_MEMBRANE_REGION	31	0.420572	1.8051884	0	0.051817805
GO_PROTEIN_COMPLEX_BINDING	28	0.36254478	1.787877	0	0.051439002
GO_REGULATION_OF_CELLULAR_COMPONENT_MOVEMENT	24	0.39252102	1.7859758	0	0.05110011

ES, NES, and FDR are enrichment score, normalized ES, and false discovery rate, respectively.



through protein–protein interactions (PPI) directly (Figure 2C). Regarding the four microRNAs in the sub-network (hsa-miR-137, hsa-miR-483-5p, hsa-miR-638, and hsa-miR-127-3p) – only hsa-miR-137 was up-regulated in the 95D cell line and the other three were down-regulated in the 95D cell lines as comparing with the 95C cell lines.

We also found that the four sub-network microRNAs were not independent of each other but connected by at least one linker protein. The three down-regulated microRNAs were previously reported to be associated with tumor metastasis in recent studies, and to influence the expression of both XPO1 and ALCAM (Ma et al., 2014; Herr et al., 2017; Shi et al., 2018; Yue et al., 2018). We focused on has-miR-137, which was up-regulated in the high-metastatic (95D) cell line and increases invasion and metastasis of NSCLC cells (Chang et al., 2017) Four up-regulated membrane proteins, HGF, CTNNB1, ITGB1, and ITGA4, involved in focal adhesion pathway, were linked to has-miR-137 by two mediation genes – PXN and MET (Figure 2B). Paxillin (PXN), whose expression was negatively correlated with has-miR-137

(Jiang et al., 2018), encodes a focal adhesion-associated protein and plays an important role in signal transduction, regulation of migration, proliferation and apoptosis. MET encodes tyrosine-protein kinase Met (c-Met) which possesses tyrosine kinase activity and is a well-characterized driver of oncogenesis occurs in multiple cancers include of NSCLC (Gentile et al., 2008). In this study, we found that MET and PXN, which are regulated by has-miR-137, may affect membrane proteins through PPI. Although proto-oncogene tyrosine-protein kinase SRC was found to be down regulated in the high-metastasis (95D) cell line in this study, many other research indicated SRC was highly expressed in NSCLC (Giaccone and Zucali, 2008; Rothschild et al., 2010).

Cancer Hallmark Pathway Enrichment

We show that our TieDIE sub-network is significantly close in the pathway space. Considering the cancer hallmark pathway enrichment of all proteins involved in the TieDIE sub-network (32 differentially expressed membrane proteins and 180 linker proteins), we found the five main cancer hallmark pathway

categories were significantly enriched, including of cell cycle pathway category (p -value = 0.0154), inflammatory response pathway category (p -value = 0.0105), metabolism pathway category (p -value = 0.0196), migration and invasion pathway category (p -value = $5.8380e-06$), and PI3K/AKT mTOR pathway category (p -value = 0.0292) (Table 2).

In addition, 10 cancer related hallmark pathways were also significantly enriched, including apoptosis pathway (p -value = 0.0221, cholesterol homeostasis pathway (p -value = 0.04152), epithelial/mesenchymal transition pathway (p -value = 0.0048), estrogen response early and late pathway (p -value = 0.04203), G2M checkpoint pathway (p -value = 0.0322), glycolysis pathway (p -value = 0.0048), IL6/Jak/Stat3 signaling pathway (p -value = 0.0067), mitotic spindle pathway (p -value = 0.0130), Mtorc1 signaling pathway (p -value = 0.0131) and TGF beta signaling pathway (p -value = 0.0500). Hallmark “wheels” were colored proportionally to the negative log transformed p -values returned by the hypergeometric test (Supplementary Table S7) (Figure 2D). These results further demonstrate the importance of our selected membrane proteins.

Prognostic Value of Differentially Expressed Membrane Proteins in TieDIE Sub-Network (Overall Survival)

The OS prognostic value of the 32 membrane proteins in the TieDIE sub-network was evaluated by performing log-rank test in the TCGA lung cancer cohort (Supplementary Table S8). Based on the best cut-off of each gene, Kaplan–Meier (KM) curves were generated for the high expression level group and the low expression level group. The expression group definition is described in section “Survival Analysis.”

High expression of CDH2 (HR = 1.2874, 95% CI: 1.0425–1.5898, p = 0.0279), EGFR (HR = 1.2927, 95% CI: 1.0496–1.5921, p = 0.0166), ITGA3 (HR = 1.2965, 95% CI: 1.0202–1.6477, p = 0.0379), ITGB1 (HR = 1.5930, 95% CI: 1.2906–1.9663, p = $2.1922e-05$), and ITGA5 (HR = 1.4656, 95% CI: 1.1836–1.8148, p = $5.8933e-04$) was negatively

associated with OS in NSCLC patients. Low expression of CALR (HR = 0.7930, 95% CI: 0.6454–0.9744, p = 0.0279) was associated with worse OS for NSCLC patients (Figure 3). According to the differential expression analysis in this work; CDH2, EGFR, ITGA3, ITGA5, and ITGB1 were all highly expressed in the high- metastatic cell lines while CALR has low expression (see Supplementary Figure S1). This is consistent with the widely accepted fact regarding NSCLC – that patients with metastasis have a poor prognosis.

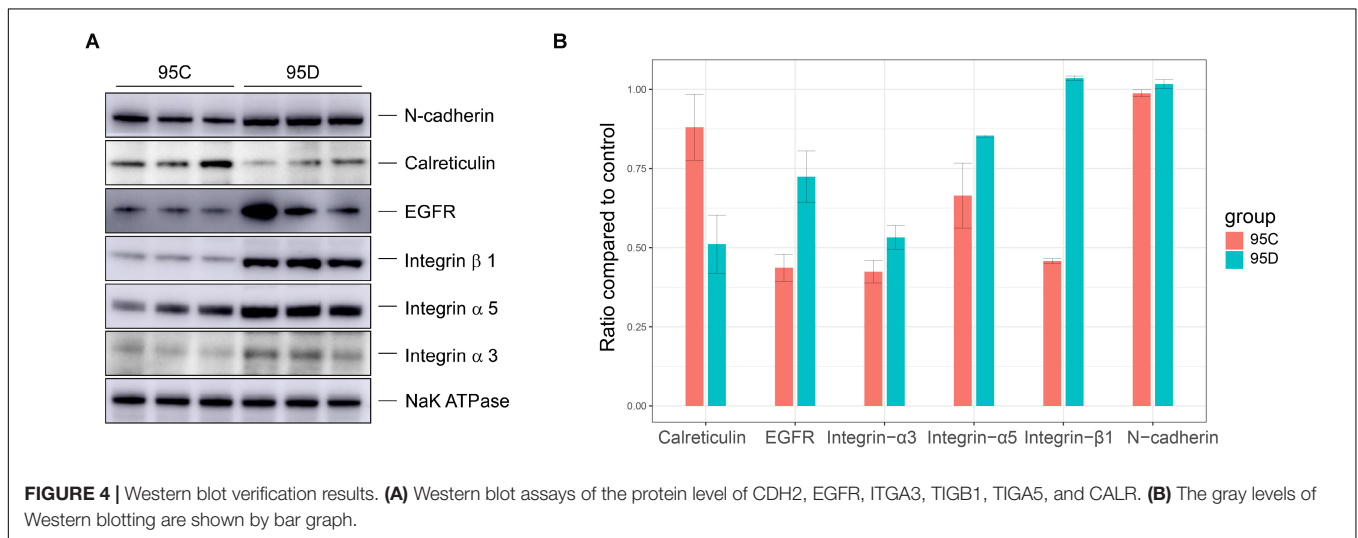
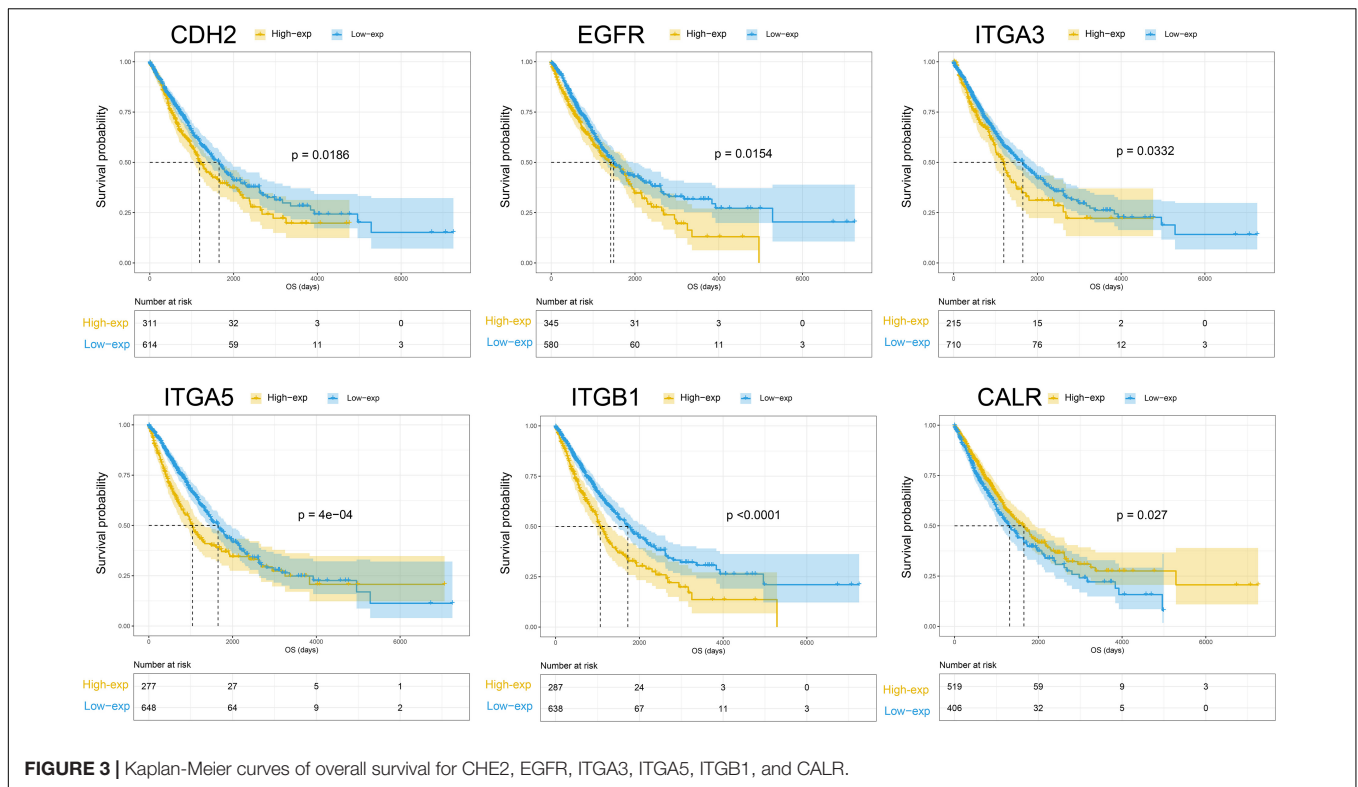
A Risk Index was computed for each sample by applying the formula described in the section “Materials and Methods.” When comparing the prognosis differences between the high and low risk factor groups, we found that the high risk group showed poor prognosis (p -value < $1e-05$) (see Supplementary Figure S2). However, not all prognosis-associated genes match this pattern. We also found five other proteins – CD47, FN1, VDAC2, HGF, and ITGA4, which showed significant correlation with OS of NSCLC patients, however, the direction of the correlation was non-intuitive. HGF, ITGA4, and CD47 were also associated with increase in OS and we can observe that patients may live longer when these genes are highly expressed. However, high expression of HGF, ITGA4, and CD47 is observed in 95D cell lines (high metastasis cell lines) as compared with low metastasis cell lines. Finally, FN1 and VDAC2 are down regulated in the 95D cell lines but lower expression level of these proteins will result in longer OS (Supplementary Figure S1 – KM curve). A potential mechanistic explanation is due to inconsistent expression of protein level and gene level (Vogel and Marcotte, 2013).

Validation of Altered Protein Expression

To validate the quantitative differences observed by mass spectrometry and bioinformatics analysis, the expression levels of six proteins were verified in the two (95C and 95D) cell lines through Western blot. The experimental results confirmed our prediction (Figure 4). Calreticulin (CALR gene production) showed low expression levels in high-metastasis cell lines (95D), while the other five proteins (CDH2, EGFR, ITGA3, ITGB1, and ITGA5) were highly expressed. The high expression of CDH2,

TABLE 2 | Cancer hallmark pathway enrichment result of all proteins involved in the TieDIE sub-network.

Hallmark main category	Hallmark sub-categories	p -value	Enrichment ratio	Gene number
Cell cycle	E2F Targets, G2M Checkpoint, p53 Pathway, Mitotic Spindle, Apoptosis	0.015443819	1.35328	29
DNA repair	PI3K/AKT/mTOR signaling, MTORC1 signaling	0.914616875	0.584921	7
Inflammatory response	IL2/STAT5 signaling, IL6/JAK/STAT3 signaling, inflammatory response, interferon alpha response, interferon gamma response, TNF alpha signaling via NFKB	0.010537424	1.56327	19
Metabolism	Hedgehog signaling, Myc targets, notch signaling, TGF beta signaling, WNT beta-catenin signaling	0.019642139	1.4941	18
Migration and invasion	Apical junction, epithelial/mesenchymal transition	$5.84E-06$	2.35103	24
Nuclear receptor signaling	DNA repair, UV response	0.199991976	1.16037	12
PI3K AKT mTOR	cholesterol homeostasis, fatty acid metabolism, glycolysis, reactive oxygen species pathway	0.029191187	1.58529	12
Stemness	Angiogenesis, hypoxia	0.302419603	1.06367	11
Tumor microenvironment	Androgen response, estrogen response early and late (merged to become nuclear receptor response)	0.401740261	0.966142	6



EGFR, ITGA3, ITGA5, and ITGB1 and the low expression of CALR in lung cancer have been validated by experimental techniques (Western blot and/or immunohistochemistry) in previous studies (Boelens et al., 2007; Grinberg-rashi et al., 2009; Wang H. et al., 2013; Liu et al., 2015; Wu et al., 2016; Du et al., 2017). Combined with the results of this study, we can infer that from the onset of lung cancer, the high expression level of the five proteins (CDH2, EGFR, ITGA3, ITGB1, and ITGA5) and the low expression of ACALR may serve as biomarkers to determine whether the tumor has a high metastasis potential.

Besides biochemical validation of the six proteins (CDH2, EGFR, ITGA3, ITGB1, and ITGA5), we also performed literature verification. *N*-cadherin (CDH2), which is a member of the cadherin family and is involved in EMT and cancer metastasis (Hazan et al., 2004), has been reported as being highly expressed in LAC tissues. It was further revealed that LAC migration and invasion are suppressed after knocking down CDH2 (Zhao et al., 2013). Regarding epidermal growth factor receptor (EGFR), its gene amplification was shown as significantly increased in tumor cells and it was closely related to metastasis and TNM stage (Jia et al., 2015). Previous studies have verified the prognostic

value of ITGA3, ITGA5, and ITGB1 expression on relapse and metastasis in lung cancer (Zheng et al., 2016). In addition, CALR was shown to be an independent prognostic factor for lung cancer (Liu et al., 2012), and the reduction in the expression of CALR was associated with an increased rate of proliferation (Bergner et al., 2009).

CONCLUSION

The high metastatic status of giant cell lung cancer is strongly associated with the abnormal expression of membrane proteins, and microRNAs play a key role in regulation of expression. From this study, we conclude that the high expression of has-miR-137 and its indirect targets-CDH2, EGFT, ITGA3, ITGB1, ITGA5 and the low expression of CALR serve as markers of high-metastasis status of giant cell lung cancer. Our study provides a new approach to the analysis of integrated proteome and microRNAs and the synthesized sub-network provides candidate targets for giant-cell lung metastasis treatment.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data generated for this study was deposited into the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) (<https://www.ebi.ac.uk/pride/>) partner repository with the dataset identifier <http://www.ebi.ac.uk/pride/archive/projects/PXD016912>.

AUTHOR CONTRIBUTIONS

YK participated in the data acquisition, performed the statistical analysis, and drafted the manuscript. ZQ and YR conceived of the study and participated in its design. GG contributed to the data

interpretation and manuscript writing. MG conducted western-blot experiment and summarized results. All authors read and approved the final manuscript. All aspects of the study were supervised by HX, HL, and HZ.

FUNDING

This work is supported by National Key R&D Program of China 2018YFC0910500, SJTU-Yale Collaborative Research Seed Fund, and the Neil Shen's SJTU Medical Research Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.01023/full#supplementary-material>

FIGURE S1 | The KM curve of 5 genes on TCGA lung cancer cohort.

FIGURE S2 | The KM curve of risk index on TCGA lung cancer cohort.

TABLE S1 | Identification results of membrane proteins.

TABLE S2 | Different expression analysis results between high and low metastatic cell lines.

TABLE S3 | Validated microRNA-regulated targets network.

TABLE S4 | GO and KEGG pathway the enrichment results.

TABLE S5 | Gene set enrichment analysis results.

TABLE S6 | Compact subnetwork predicted by TieDIE program.

TABLE S7 | cancer hallmark pathway enrichment analysis of genes involved in subnetwork.

TABLE S8 | Raw clinical data and 32 membrane protein expression data on TCGA lung cancer cohort.

REFERENCES

- Arend, R. C., Londono-Joshi, A. I., Michael Straughn, J., and Buchsbaum, D. J. (2013). Gynecologic oncology the Wnt/ β -catenin pathway in ovarian cancer: a review. *Gynecol. Oncol.* 131, 772–779. doi: 10.1016/j.ygyno.2013.09.034
- Benjamini, Y., and Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Stat.* 24, 407–418. doi: 10.1111/1467-9469.00072
- Bergner, A., Kellner, J., Tufman, A., and Huber, R. M. (2009). Endoplasmic reticulum Ca^{2+} -homeostasis is altered in small and non-small cell lung cancer cell lines. *J. Exp. Clin. Cancer Res.* 28, 1–7. doi: 10.1186/1756-9966-28-25
- Berrout, J., Kyriakopoulou, E., Moparthy, L., Hoge, A. S., Berrout, L., Ivan, C., et al. (2017). TRPA1-FGFR2 binding event is a regulatory oncogenic driver modulated by miRNA-142-3p. *Nat. Commun.* 8, 947–962. doi: 10.1038/s41467-017-00983-w
- Bland, J. M., and Altman, D. G. (1998). Survival probabilities (the Kaplan-Meier method). *BMJ* 317, 1572–1580. doi: 10.1136/bmj.317.7172.1572
- Boelens, M. C., Van Den Berg, A., Vogelzang, L., Wesseling, J., Postma, D. S., Timens, W., et al. (2007). Differential expression and distribution of epithelial adhesion molecules in non-small cell lung cancer and normal bronchus. *J. Clin. Pathol.* 60, 608–614. doi: 10.1136/jcp.2005.031443
- Borghaei, H., Pazares, L., Horn, L., Spigel, D. R., Steins, M., Ready, N. E., et al. (2017). Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* 373, 1627–1639. doi: 10.1056/NEJMoa1507643. Nivolumab
- Brown, A., Mortier, R., and Rodden, T. (2013). MultiNet: Reducing interaction overhead in domestic wireless networks. *Hum. Factors Comput. Syst.* 1569–1578. doi: 10.1145/2470654.2466208
- Chang, T. H., Tsai, M. F., Gow, C. H., Wu, S. G., Liu, Y. N., Chang, Y. L., et al. (2017). Upregulation of microRNA-137 expression by Slug promotes tumor invasion and metastasis of non-small cell lung cancer cells through suppression of TFAP2C. *Cancer Lett.* 402, 190–202. doi: 10.1016/j.canlet.2017.06.002
- Chen, W. Y., Tsai, Y. C., Siu, M. K., Yeh, H. L., Chen, C. L., Yin, J. J., et al. (2017). Inhibition of the androgen receptor induces a novel tumor promoter, ZBTB46, for prostate cancer metastasis. *Oncogene* 36, 6213–6224. doi: 10.1038/onc.2017.226
- de Wit, M., Jimenez, C. R., Carvalho, B., Belien, J. A. M., Diemen, P. M. D., Mongera, S., et al. (2012). Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. *Gut* 61, 855–864. doi: 10.1136/gutjnl-2011-300511
- Ding, X., Yang, Y., Han, B., Du, C., Xu, N., Huang, H., et al. (2014). Transcriptomic characterization of hepatocellular carcinoma with CTNNB1 mutation. *PLoS One* 9:e95307. doi: 10.1371/journal.pone.0095307
- Drake, J. M., Paull, E. O., Graham, N. A., Lee, J. K., Smith, B. A., Titz, B., et al. (2016). Phosphoproteome integration reveals patient-specific networks in prostate cancer. *Cell* 166, 1041–1054. doi: 10.1016/j.cell.2016.07.007

- Du, H., Chen, Y., Hou, X., Huang, Y., Wei, X., Yu, X., et al. (2017). PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. *Cell Death Dis.* 8:e3143. doi: 10.1038/cddis.2017.553
- Dweep, H., and Gretz, N. (2015). miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods* 12, 697–697. doi: 10.1038/nmeth.3485
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207
- Ge, M., Qiao, Z., Kong, Y., Lu, H., and Liu, H. (2020). Exosomes mediate intercellular transfer of non-autonomous tolerance to proteasome inhibitors in mixed-lineage leukemia. *Cancer Sci.* 111, 1–12. doi: 10.1111/cas.14351
- Gentile, A., Trusolino, L., and Comoglio, P. M. (2008). The Met tyrosine kinase receptor in development and cancer. *Cancer Metastasis Rev.* 27, 85–94. doi: 10.1007/s10555-007-9107-6
- Giaccone, G., and Zucali, P. A. (2008). Src as a potential therapeutic target in non-small-cell lung cancer. *Ann. Oncol.* 19, 1219–1223. doi: 10.1093/annonc/mdn048
- Grinberg-rashi, H., Ofek, E., Perelman, M., Skarda, J., Jacob-hirsch, J., Amariglio, N., et al. (2009). The expression of three genes in primary non-small cell lung cancer is associated with metastatic spread to the brain. *Clin. Cancer Res.* 15, 1755–1762. doi: 10.1158/1078-0432.CCR-08-2124
- Gumbiner, B. M. (1996). Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* 84, 345–357. doi: 10.1016/S0092-8674(00)81279-9
- Hazan, R. B., Qqiao, R., Keren, R., Badano, I., and Suyama, K. (2004). Cadherin switch in tumor progression. *Ann. N.Y. Acad. Sci.* 1014, 155–163. doi: 10.1196/annals.1294.016
- Herr, I., Sähr, H., Zhao, Z., Yin, L., Omlor, G., Lehner, B., et al. (2017). MiR-127 and miR-376a act as tumor suppressors by in vivo targeting of COA1 and PDIA6 in giant cell tumor of bone. *Cancer Lett.* 409, 49–55. doi: 10.1016/j.canlet.2017.08.029
- Inamura, K., and Ishikawa, Y. (2016). MicroRNA in lung cancer: novel biomarkers and potential tools for treatment. *J. Clin. Med.* 5, 36–49. doi: 10.3390/jcm5030036
- Iqbal, M. A., Arora, S., Prakasam, G., Calin, G. A., and Syed, M. A. (2018). MicroRNA in lung cancer: role, mechanisms, pathways and therapeutic relevance. *Mol. Aspects Med.* 70, 3–20. doi: 10.1016/j.mam.2018.07.003
- Jia, X. F., Li, J., Zhao, H. B., Liu, J., and Liu, J. J. (2015). Correlation of EGFR gene amplification with invasion and metastasis of non-small cell lung cancer. *Genet. Mol. Res.* 14, 11006–11012. doi: 10.4238/2015.September.21.13
- Jiang, H., Zhang, H., Hu, X., and Li, W. (2018). Knockdown of long non-coding RNA XIST inhibits cell viability and invasion by regulating miR-137/PXN axis in non-small cell lung cancer. *Int. J. Biol. Macromol.* 111, 623–631. doi: 10.1016/j.ijbiomac.2018.01.022
- Jiang, W. G., Sanders, A. J., Katoh, M., Ungefroren, H., Gieseler, F., Prince, M., et al. (2015). Seminars in cancer biology tissue invasion and metastasis: molecular, biological and clinical perspectives. *Semin. Cancer Biol.* 35, 244–275. doi: 10.1016/j.semcancer.2015.03.008
- Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., et al. (2019). Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* 567, 257–261. doi: 10.1038/s41586-019-0987-8
- Jungo, F., Bougueleret, L., Xenarios, I., and Poux, S. (2012). Toxicon the UniProtKB/Swiss-Prot Tox-Pro program: a central hub of integrated venom protein data. *Toxicon* 60, 551–557. doi: 10.1016/j.toxicon.2012.03.010
- Kampen, K. (2011). Membrane proteins: the key players of a cancer cell. *J. Membr. Biol.* 242, 69–74. doi: 10.1007/s00232-011-9381-7
- Kelstrup, C. D., Young, C., Lavallee, R., Nielsen, M. L., and Olsen, J. V. (2012). Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* 11, 3487–3497. doi: 10.1021/pr3000249
- Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. *PLoS Comput. Biol.* 9:e1002886. doi: 10.1371/journal.pcbi.1002886
- Kramarz, B., Huntley, R. P., Rodríguez-López, M., Roncaglia, P., Saverimuttu, S. C. C., Parkinson, H., et al. (2020). Gene ontology curation of neuroinflammation biology improves the interpretation of Alzheimer's disease gene expression data. *J. Alzheimer's Dis.* 75, 1417–1435. doi: 10.3233/jad-200207
- Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Latouche, A., and Aurelien, A. (2019). CRAN Task View: Survival Analysis. Available online at: <http://cran.r-project.org/view=Survival>
- Le, N. Q. K., Ho, Q. T., and Ou, Y. Y. (2017a). Incorporating deep learning with convolutional neural networks and position specific scoring matrices for identifying electron transport proteins. *J. Comput. Chem.* 38, 2000–2006. doi: 10.1002/jcc.24842
- Le, N. Q. K., Nguyen, T. T. D., and Ou, Y. Y. (2017b). Identifying the molecular functions of electron transport proteins using radial basis function networks and biochemical properties. *J. Mol. Graph. Model.* 73, 166–178. doi: 10.1016/j.jmkgm.2017.01.003
- Le, N. Q. K., Yapp, E. K. Y., Nagasundaram, N., Chua, M. C. H., and Yeh, H. Y. (2019a). Computational identification of vesicular transport proteins from sequences using deep gated recurrent units architecture. *Comput. Struct. Biotechnol. J.* 17, 1245–1254. doi: 10.1016/j.csbj.2019.09.005
- Le, N. Q. K., Yapp, E. K. Y., and Yeh, H. Y. (2019b). ET-GRU: using multi-layer gated recurrent units to identify electron transport proteins. *BMC Bioinformatics* 20:377. doi: 10.1186/s12859-019-2972-5
- Lei, C., Du, F., Sun, L., Li, T., Li, T., Min, Y., et al. (2017). miR-143 and miR-145 inhibit gastric cancer cell migration and metastasis by suppressing MYO6. *Cell Death Dis.* 8:e3101. doi: 10.1038/cddis.2017.493
- Lethlarsen, R., Raaen Lund, R., Hansen, H. V., Laenkholt, A., Tarin, D., Jensen, O. N., et al. (2009). Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. *Mol. Cell. Proteomics* 8, 1436–1449. doi: 10.1074/mcp.M800061-MCP200
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi: 10.1093/bioinformatics/btr260
- Liu, C., Wang, X., Genchev, G. Z., and Lu, H. (2017). Multi-omics facilitated variable selection in Cox-regression model for cancer prognosis prediction. *Methods* 124, 100–107. doi: 10.1016/j.ymeth.2017.06.010
- Liu, R., Gong, J., Chen, J., Li, Q., Song, C., Zhang, J., et al. (2012). Calreticulin as a potential diagnostic biomarker for lung cancer. *Cancer Immunol. Immunother.* 61, 855–864. doi: 10.1007/s00262-011-1146-8
- Liu, X., Sun, N., Dong, Y., Li, J., Liu, Y., Ren, Y., et al. (2015). Anticancer effects of adenovirus-mediated calreticulin and melanoma-associated antigen 3 expression on non-small cell lung cancer cells. *Int. Immunopharmacol.* 25, 416–424. doi: 10.1016/j.intimp.2015.02.017
- Ma, K., Pan, X., Fan, P., He, Y., Gu, J., Wang, W., et al. (2014). Loss of miR-638 in vitro promotes cell invasion and a mesenchymal-like transition by influencing SOX2 expression in colorectal carcinoma cells. *Mol. Cancer* 13, 1–13. doi: 10.1186/1476-4598-13-118
- Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., et al. (2011). Mass spectrometry-based proteomics using Q exactive, a high-performance benchtop quadrupole orbitrap mass spectrometer. *Mol. Cell. Proteomics* 10, 1–11. doi: 10.1074/mcp.M111.011015
- Mo, D., Yang, D., Xiao, X., Sun, R., Huang, L., and Xu, J. (2017). MiRNA-145 suppresses lung adenocarcinoma cell invasion and migration by targeting N-cadherin. *Biotechnol. Lett.* 39, 701–710. doi: 10.1007/s10529-017-2290-9
- Nicoloso, M. S., Spizzo, R., Shimizu, M., Rossi, S., and Calin, G. A. (2009). MicroRNAs - The micro steering wheel of tumour metastases. *Nat. Rev. Cancer* 9, 293–302. doi: 10.1038/nrc2619
- Paull, E. O., Carlin, D. E., Niepel, M., Sorger, P. K., Haussler, D., and Stuart, J. M. (2013). Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics* 29, 2757–2764. doi: 10.1093/bioinformatics/btt471
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, 442–450. doi: 10.1093/nar/gky1106
- Przybyla, P., Shardlow, M., Aubin, S., Bossy, R., De Castilho, R. E., Piperidis, S., et al. (2016). Text mining resources for the life sciences. *Database* 2016, 1–30. doi: 10.1093/database/baw145
- Qiao, Z., Zhang, Y., Ge, M., Liu, S., Jiang, X., Zhi, S., et al. (2019). Cancer cell derived small extracellular vesicles contribute to recipient cell metastasis

- through promoting HGF/c-Met pathway. *Mol. Cell. Proteomics* 18, 1619–1629. doi: 10.1074/mcp.RA119.001502
- Rothschild, S., Gautschi, O., Haura, E., and Johnson, F. (2010). Src inhibitors in lung cancer: current status and future directions. *Clin. Lung Cancer* 11, 238–242. doi: 10.3816/CLC.2010.n.030
- Rutnam, Z. J., Wight, T. N., and Yang, B. B. (2013). miRNAs regulate expression and function of extracellular matrix molecules. *Matrix Biol.* 32, 74–85. doi: 10.2217/FON.09.6.Dendritic
- Samaekia, R., Adorno-Cruz, V., Bockhorn, J., Chang, Y. F., Huang, S., Prat, A., et al. (2017). miR-206 inhibits stemness and metastasis of breast cancer by targeting MKL1/IL11 pathway. *Clin. Cancer Res.* 23, 1091–1103. doi: 10.1158/1078-0432.CCR-16-0943
- Shi, M., Jiang, Y., Yang, L., Yan, S., Wang, Y. G., and Lu, X. J. (2018). Decreased levels of serum exosomal miR-638 predict poor prognosis in hepatocellular carcinoma. *J. Cell. Biochem.* 119, 4711–4716. doi: 10.1002/jcb.26650
- Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics, 2017. *CA. Cancer J. Clin.* 67, 7–30. doi: 10.3322/caac.21387
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review the cancer genome atlas (TCGA): an immeasurable source of knowledge. *Wspolczesna Onkol. Oncol.* 2015, 68–77. doi: 10.5114/wo.2014.47136
- Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., et al. (2010). Towards a knowledge-based human protein atlas. *Nat. Biotechnol.* 28, 1248–1250. doi: 10.1038/nbt1210-1248
- Vadakekolathu, J., Al-Juboori, S. I. K., Johnson, C., Schneider, A., Buczek, M. E., Di Biase, A., et al. (2018). MTSS1 and SCAMP1 cooperate to prevent invasion in breast cancer. *Cell Death Dis.* 9, 344–344. doi: 10.1038/s41419-018-0364-9
- Vogel, C., and Marcotte, E. M. (2013). Insights into regulation of protein abundance from proteomics and transcriptomics analyses. *Nat. Rev. Genet.* 13, 227–232. doi: 10.1038/nrg3185
- Wang, H., Zhu, Y., Zhao, M., Wu, C., Zhang, P., Tang, L., et al. (2013). miRNA-29c suppresses lung cancer cell adhesion to extracellular matrix and metastasis by targeting integrin $\beta 1$ and matrix metalloproteinase 2 (MMP2). *PLoS One* 8:e70192. doi: 10.1371/journal.pone.0070192
- Wang, X. M., Li, J., Yan, M. X., Liu, L., Jia, D. S., Geng, Q., et al. (2013). Integrative analyses identify osteopontin, LAMB3 and ITGB1 as critical Pro-metastatic genes for lung cancer. *PLoS One* 8:e55714. doi: 10.1371/journal.pone.0055714
- Waqar, S. N., Samson, P. P., Robinson, C. G., Bradley, J., Devarakonda, S., Du, L., et al. (2018). Non-small-cell lung cancer with brain metastasis at presentation. *Clin. Lung Cancer* 19, e373–e379. doi: 10.1016/j.clcc.2018.01.007
- Wu, D. W., Chen, T. C., Huang, H. S., and Lee, H. (2016). TC-N19, a novel dual inhibitor of EGFR and cMET, efficiently overcomes EGFR-TKI resistance in non-small-cell lung cancer cells. *Cell Death Dis.* 7:e2290. doi: 10.1038/cddis.2016.192
- Wu, G., Maharjan, S., Kim, D., Kim, J. N., Park, B. K., Koh, H., et al. (2018). A novel monoclonal antibody targets Mucin1 and attenuates growth in pancreatic cancer model. *Int. J. Mol. Sci.* 19:2004. doi: 10.3390/ijms19072004
- Xu, J., Zhang, C., Wang, X., Zhai, L., Ma, Y., Mao, Y., et al. (2020). Integrative proteomic characterization of human lung adenocarcinoma. *Cell* 182, 245–261.e17. doi: 10.1016/j.cell.2020.05.043
- Ye, P., Shi, Y., An, N., Zhou, Q., Guo, J., and Long, X. (2018). miR-145 overexpression triggers alteration of the whole transcriptome and inhibits breast cancer development. *Biomed. Pharmacother.* 100, 72–82. doi: 10.1016/j.biopha.2018.01.167
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. A J. Intergrative Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yue, J., Lv, D., Wang, C., Li, L., Zhao, Q., Chen, H., et al. (2018). Epigenetic silencing of miR-483-3p promotes acquired gefitinib resistance and EMT in EGFR-mutant NSCLC by targeting integrin $\beta 3$. *Oncogene* 37, 4300–4312. doi: 10.1038/s41388-018-0276-2
- Zhang, L., Cai, J., Fang, L., Huang, Y., Li, R., Xu, X., et al. (2017). Simultaneous overactivation of Wnt/ β -catenin and TGF β signalling by miR-128-3p confers chemoresistance-associated metastasis in NSCLC. *Nat. Commun.* 8, 1–18. doi: 10.1038/ncomms15870
- Zhang, S., Ge, W., Zou, G., Yu, L., Zhu, Y., Li, Q., et al. (2018). MiR-382 targets GOLM1 to inhibit metastasis of hepatocellular carcinoma and its down-regulation predicts a poor survival. *Am. J. Cancer Res.* 8, 120–131.
- Zhao, J. Q., Sun, F. J., Liu, S. S., Yang, J., Wu, Y. Q., Li, G. S., et al. (2013). Expression of connexin 43 and E-cadherin protein and mRNA in non-small cell lung cancers in Chinese patients. *Asian Pacific J. Cancer Prev.* 14, 639–643. doi: 10.7314/APJCP.2013.14.2.639
- Zheng, W., Jiang, C., and Li, R. (2016). Integrin and gene network analysis reveals that ITGA5 and ITGB1 are prognostic in non-small-cell lung cancer. *Onco. Targets. Ther.* 9, 2317–2327. doi: 10.2147/OTT.S91796

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kong, Qiao, Ren, Genchev, Ge, Xiao, Zhao and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.