RESEARCH ARTICLE
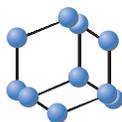
Single-cell RNA Sequencing Analysis Identifies Key Genes in Brain Metastasis from Lung Adenocarcinoma

Zilong Zhang[1,2], Feifei Cui[1,2], Murong Zhou[3], Song Wu[4,*], Quan Zou[1,2,5,*] and Bo Gao[6,*]

[1]*Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China;* [2]*Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China;* [3]*College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China;* [4]*Director of Preventive Treatment of Disease Centre, Qinhuangdao Hospital of Traditional Chinese Medicine, Qinhuangdao 066000, China;* [5]*Hainan Key Laboratory for Computational Science and Application, Hainan Normal University, Haikou571158, China;* [6]*Second Affiliated Hospital of Harbin Medical University, Harbin Medical University, Harbin, 150080, China*

Abstract: ***Background***: Lung adenocarcinoma (LADC) is the most common type of lung cancer and is a subtype of non-small-cell lung cancer (NSCLC). Approximately 40% of LADC patients experience brain metastases (BMs) during the course of the disease. In this study, integrated bioinformatics methods were applied to identify key genes related to brain metastasis in lung adenocarcinoma.

Methods: We derived and characterized genes differentially expressed between the primary tumour and brain metastases using tumour cells isolated from two lung cancer Patient-derived xenografts (PDX) cases (GSE 69405). Gene ontology (GO) and KEGG pathway enrichment analyses were applied, and protein-protein interaction (PPI) networks and Cytoscape software were utilized to identify key genes.

Results: Four key genes, including *CKAP4 (Cytoskeleton Associated Protein 4)*, *SERPINA1 (Serpin Family A Member 1)*, *SDC2 (Syndecan 2)* and *GNG11 (G Protein Subunit Gamma 11)* were identified for BM-LADC by the Venn diagram.

Conclusion: We believe these key genes may be potential biomarkers for improved prognosis and treatment of lung adenocarcinoma.

1. INTRODUCTION

Lung adenocarcinoma (LADC) is the most common type of lung cancer and type of non-small-cell lung cancer (NS-CLC) [1]. Approximately 40% of LADC patients experience brain metastases (BMs) during the course of the disease [2, 3]. There are few treatment options for BM-LADC, mainly including surgery and radiosurgery [4]. Moreover, these treatments are usually ineffective, leading to a low survival rate [5]. Therefore, there is an urgent need to uncover the key genes and signalling pathways to reduce BM by early diagnosis.

Intensive studies have focused on understanding the complex process of BM-LADC [6-9]. Shih *et al.* confirmed that overexpression of *MYC*, *YAP1* and *MMP13* can increase the incidence of brain metastasis [5]. Pocha *et al.* defined a subtype of brain metastasis from lung adenocarcinoma with the expression of *SFTPA1, SFTPB* and *NAPSA* [10]. However, the transition mechanisms between primary tumours and brain metastases are still not quite clear. Biomarkers for identifying "premetastatic" lesions would be useful in diagnosing BM and providing actionable targets [11]. In the past, RNA sequencing has mostly been conducted in tissue, which is also known as bulk-seq [12]. Due to the complex tumour microenvironment, sequencing the average gene expression only in tissue is not sufficient for understanding the disease [13, 14].

Single-cell RNA sequencing (scRNA-seq), a relatively new technique, has been widely used to identify therapeutic targets and biomarkers in many diseases [15-26]. Unlike traditional bulk RNA sequencing, scRNA-seq quantifies the gene expression for each single cell [27, 28]; consequently, scRNA-seq is especially useful to uncover complex tumour tissues that contain cells of different types and cancer stages [29-32]. Similarly, scRNA-seq could also uncover the trajectory of dynamic changes in cell state [33], and hence is of

* Address correspondence to these authors at the Director of Preventive Treatment of Disease Centre, Qinhuangdao Hospital of Traditional Chinese Medicine, Qinhuangdao 066000, China; E-mail: drwusong616@sina.com; Hainan Key Laboratory for Computational Science and Application, Hainan Normal University, Haikou 571158, China;
E-mail: zouquan@nclab.net; Second Affiliated Hospital of Harbin Medical University, Harbin Medical University, Harbin, 150080, China;
E-mail: 1678729588@qq.com

great use to identify key genes in tumour metastasis [34, 35].

In this study, integrated bioinformatics methods were applied to identify key genes in brain metastases from lung adenocarcinoma. We derived and characterized genes differentially expressed between the primary tumour and brain metastasis using tumour cells isolated from two lung cancer PDX cases (GSE 69405) [36]. Gene ontology (GO) and KEGG pathway enrichment analyses were applied, and protein-protein interaction (PPI) networks and Cytoscape software were utilized to identify key genes [37-42]. Four key genes, *CKAP4, SERPINA1, SDC2,* and *GNG11*, were identified for BM-LADC by Venn diagram. We believe these key genes may be potential biomarkers for better prognosis and treatment of lung adenocarcinoma.

2. MATERIALS AND METHODS

2.1. Acquisition of LADC Cell Samples

We obtained LADC cell samples from the publicly accessible Gene Expression Omnibus (GEO) database by downloading the transcriptome profile from GSE69405. An expression matrix consisting of 128 cancer cells was obtained by combining tumour cell-enriched PDX cells (LC-PT-45), an additional PDX cell batch (LC-PT-45) and another lung cancer brain metastasis PDX case (LC-MBT-15).

2.2. Processing scRNA-seq Data

We chose the popular single-cell analysis tool Seurat package to process the data [43]. Following the standard process of Seurat, low-quality single cells were first filtered using a number of detected genes and percentage of mitochondria sequencing count. In this case, cells with fewer than 6,000 detected genes and a percentage of mitochondria sequencing count higher than 35% were excluded as low-quality cells. In addition, we followed the standard Seurat pipeline and calculated the most variable genes. Then, linear dimension reduction PCA and nonlinear dimension reduction method t-SNE were utilized, and both visualization results showed two separate clusters, indicating that significant differences existed between the primary tumour cells and brain metastasis tumour cells. Afterward, the top 500 differentially expressed genes between these two clusters were identified as marker genes.

2.3. Enrichment Analysis

Gene Ontology (GO) annotations and KEGG pathway analysis were performed using the R package clusterProfiler (version 3.16.1) [44-46]. GO terms were divided into three groups: biological processed (BP), cellular components (C-C) and molecular functions (MF) [47]. KEGG pathway enrichment analysis was also carried out using the R package clusterProfiler (version 3.16.1). Group p-values were set to lower than 0.05, and the miminum size of genes in each group was set to 10. For the enrichment of hub genes, we chose the DAVID (Database for Annotation, Visualization, and Integrated Discovery) database [48].

2.4. PPI Network Analysis

The STRING database (Version 11.0) was utilized to evaluate interactions between proteins (https://string-db.org/) using the top 500 DEGs [49]. We screened important interactions by considering a combined constructed score lower than 0.9 to be significant. The PPI results were downloaded and further analysed using Cytoscape (version 3.7.2) software [50].

A Cytoscape plug-in, "Molecular Complex Detection" (MCODE) [51], was utilized to screen significant PPI network modules. The parameters were set with a degree cut-off of 2, node score cut-off of 0.2, k-core of 2, and a maximum depth of 100. Another Cytoscape plug-in, Cytohubba [52] was used to find the hub genes in the PPI network. The maximal clique size (MCC) was used to calculate the top 10 nodes, which represent the most significant hub genes.

2.5. Identification of Key Biomarkers

A Venn diagram was used to identify key biomarkers among "significant genes identified by PPI degree", "hub genes identified by Cytohubba," and "hub genes identified by MCODE". The Venn diagram was drawn using the website (http://bioinformatics.psb.ugent.be/webtools/Venn/). Functions for the four identified key genes were obtained *via* GeneCards (https://www.genecards.org/) [53].

3. RESULTS

3.1. scRNA-seq Data Profiling

We acquired 126 high-quality cells from LADC patients; among them, 77 cells were isolated in PDM from the primary tumour (pt), whereas 49 cells were isolated in PDM from brain metastasis tumours (mbt). All gene expression values downloaded from GEO (Gene Expression Omnibus) of these 126 cells were combined into a matrix. Quality control is shown in Fig. (**1A**), which illustrates the detected genes for each cell, the library size for each cell, and the percent of mitochondria counts for each cell. Based on the results in Fig. (**1A**), we excluded cells with detected genes < 6000 and mitochondrial counts > 35%, resulting in 120 cells remaining. Fig. (**1B**) shows a positive correlation (Pearson's r = 0.29) between library size and detected genes. Fig. (**1C**) illustrates the highly differentially expressed genes (DEGs) across all 120 cells, and the top 10 significantly DEGs were *SAA1, LAN2, HIST1H4C, SAA2, CD44, SAT1, GSTP1, IL8, ASPH,* and *EREG* (Table **1**). Principal component analysis (PCA) was used to visualize these cells, and the results are shown in Fig. (**1D**) prove that the gene expression levels of PT cells and MBT cells are significantly different in two separate clusters. In addition, standard deviations of the principal components were calculated, as shown in Fig. (**1E**), and the elbow indicated that the first 20 dimensions of the principal components were sufficient for further analysis. t-Distributed stochastic neighbour embedding (t-SNE) was then conducted to verify the visualization results of PCA (Fig. **1F**). As expected, pt and mbt cells were clustered into two groups. Accordingly, we performed differential analysis and

displayed the top 20 significantly differentially expressed genes *via* a heatmap (Fig. **1G**). We then characterized the trajectory of the 120 single cells (Fig. **1H**) and showed a signifi- cant tendency curve from the primary tumour (pt) to the brain metastasis tumour (mbt), indicating the possibility of uncovering key genes in the BM-LADC process.
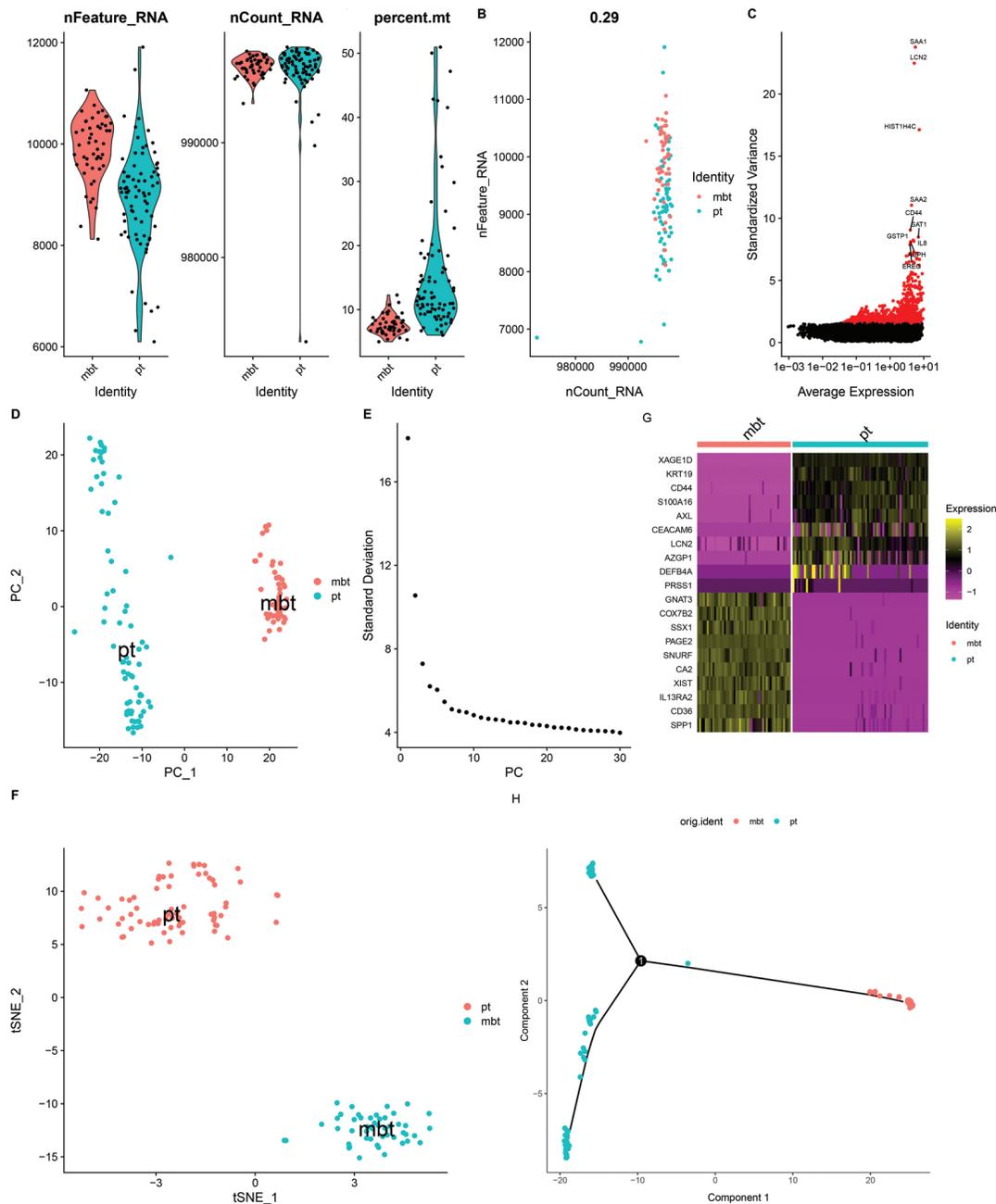

Fig. (1). Processing single-cell RNA sequencing data. (**A**) Quality control of scRNA-seq data. We filtered out low-quality cells by detected gene count, library size and percentage of mitochondrial genes. After filtering, 120 single cells were obtained for further analysis. (**B**) Pearson's correlation (r = 0.29) between detected genes and library sizes. (**C**) Volcano plot assessment of differentially expressed genes. Red dots indicate highly variable genes between pt cells and mbt cells. (**D**) Linear dimension reduction technique Principal component analysis (PCA) was applied to scRNA-seq data, and the results showed that bt cells and mbt cells were grouped into separate clusters. (**E**) Principle components were ranked based on the percentage of variance explained by each principal component (elbow plot). (**F**) The nonlinear dimension reduction method t-distributed stochastic neighbour embedding (t-SNE) was applied to scRNA-seq data. (**G**) A heatmap was constructed using the top 10 significant marker genes between pt cells and mbt cells. (H) Trajectory analysis revealed a significant tendency curve from the primary tumour (pt) to the brain metastasis tumour (mbt), indicating the possibility of uncovering key genes in the BM-LADC process. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table 1. Screening DEGs in mbt cells.

DEGs	Official Gene Symbol
Upregulated	*DSCR8, GNAT3, COX7B2, AC006050.2, SSX5, EDIL3, GTSF1, SNRPN, MAGEC2, TEX41*
Downregulated	*XAGE1D, KRT19, XAGE1B, XAGE1E, ST3GAL1, CD44, S100A16, AXL, CEACAM6, CD163L1*

Abbreviations are as follows: DEGs, differentially expressed genes.
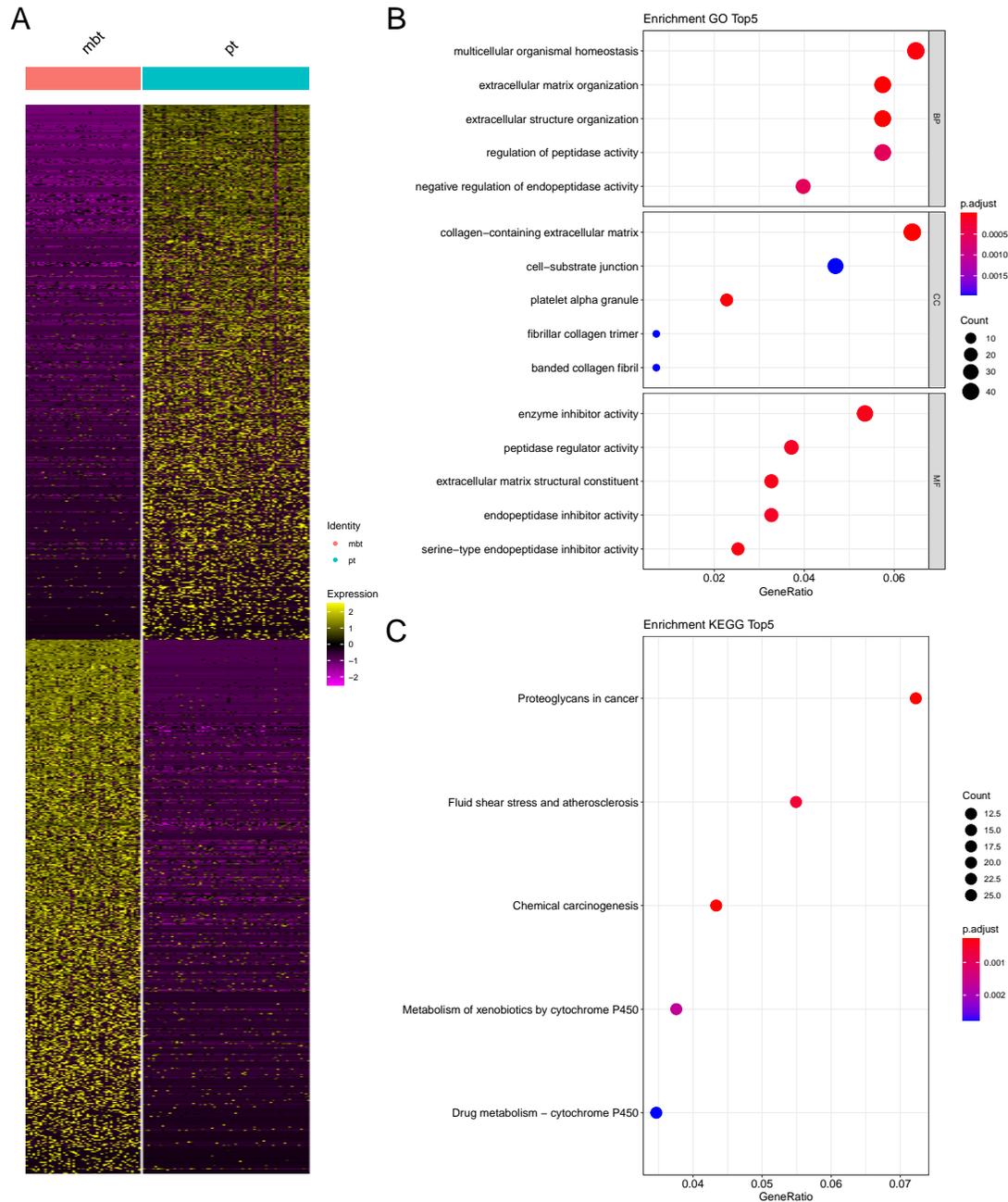

Fig. (2). Enrichment results of differentially expressed genes. (**A**) Heatmap displaying the top 500 differentially expressed genes (DEGs) per cluster. (**B**) Dot plot using the top 500 DEGs depicting the top 5 terms of GO enrichment in three categories, namely, BP, CC and MF. (**C**) Dot plot showing the top 5 significantly enriched KEGG pathways using the top 500 DEGs. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

A
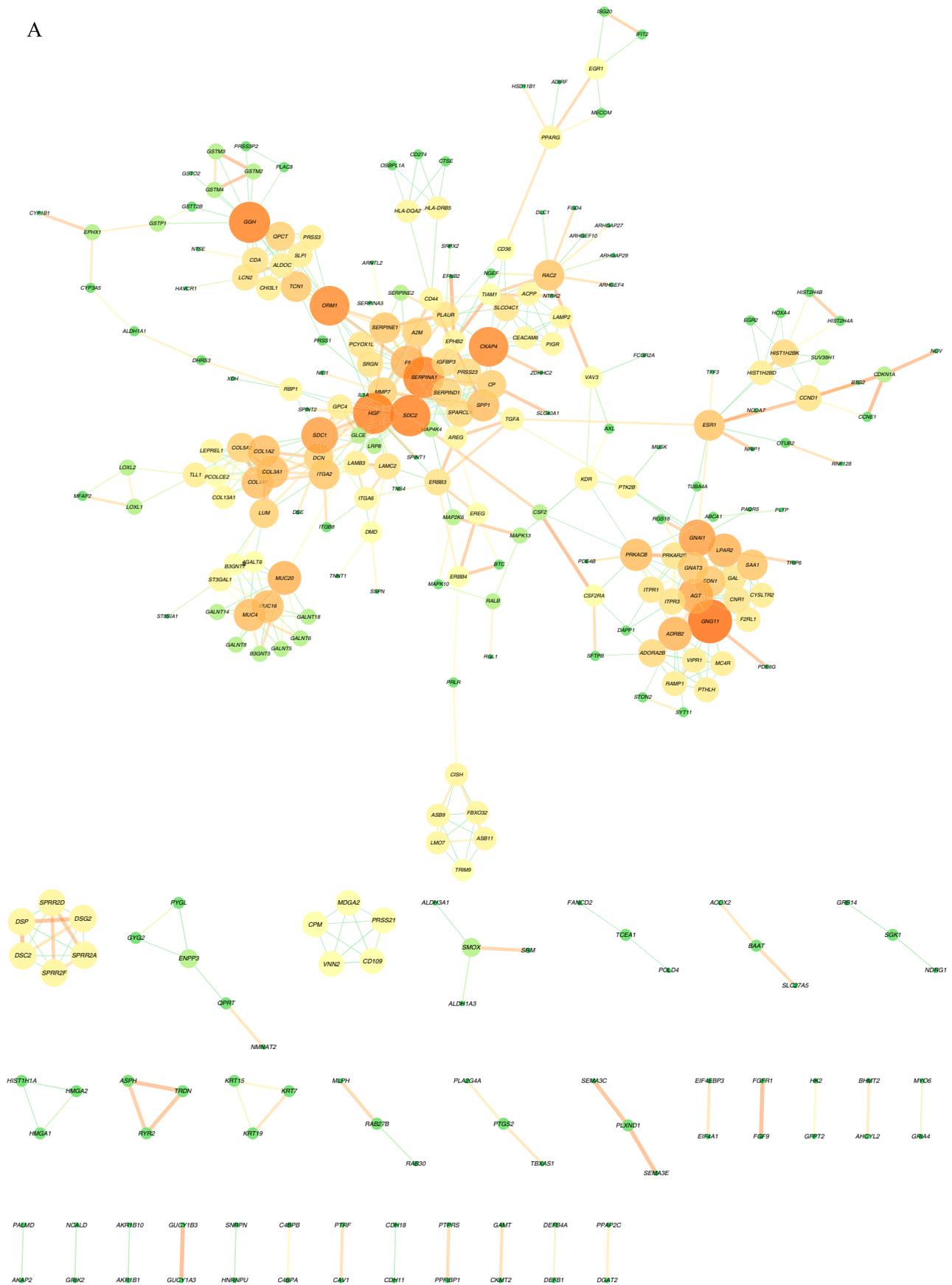
(Fig. 3) contd....
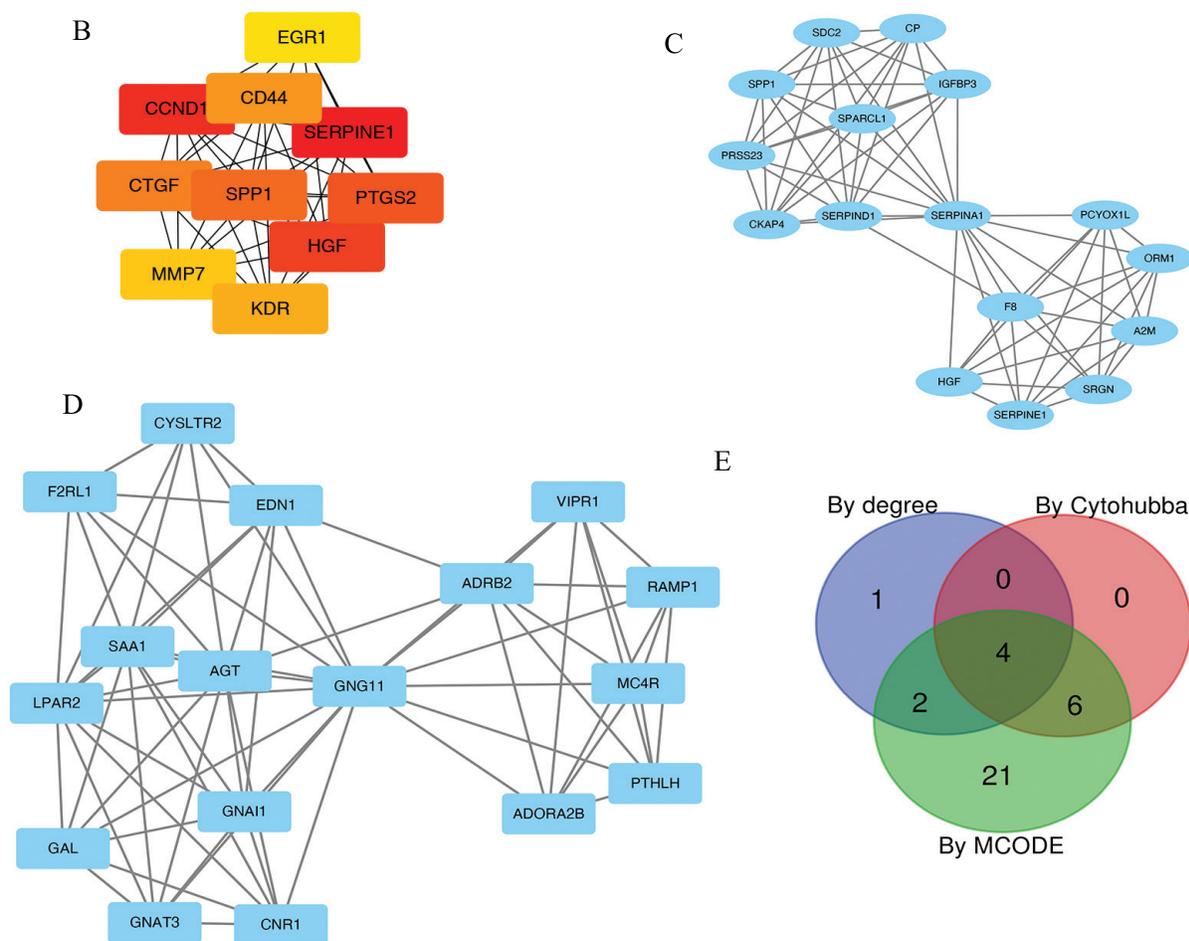
Fig. (3). PPI networks and VENN diagram. (**A**) The overall view of the PPI network constructed using the top 500 DEGs between primary LADC cells and brain metastasis tumour cells. (**B**) Hub genes identified by Cytohubba. (**C**) and (**D**) Hub genes identified by MCODE. (**E**) Venn diagram identifying key genes by finding the common genes identified by degree, CytoHubba, and MCODE. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

3.2. Enrichment Analysis

With the DEGs obtained from scRNA-seq analysis, (Fig. **2A**) functional annotation was carried out. The GO terms were divided into three categories, namely, biological processed (BP), cellular components (CC), and molecular functions (MF). The results are shown in Fig. (**2B**). For BP, the top 5 enriched GO terms were multicellular organismal homeostasis, extracellular matrix organization, extracellular structure organization, regulation of peptidase activity and negative regulation of endopeptidase activity. The GO-CC terms were mainly enriched in collagen-containing extracellular matrix, cell-substrate junction, platelet alpha granule, fibrillar collagen trimer and banded collagen fibril. The most enriched GO-MF terms were enzyme inhibitor activity, peptidase regulator activity, extracellular matrix structural constituent, endopeptidase inhibitor activity and serine-type endopeptidase inhibitor activity.

The most significantly enriched KEGG pathways are shown in Fig. (**2C**). The significant KEGG pathways included proteoglycans in cancer, fluid shear stress and atherosclerosis, chemical carcinogenesis, metabolism of xenobiotics

by cytochrome P450, and drug metabolism-cytochrome P450 (Fig. **2**).

3.3. Protein-Protein Interaction (PPI) Network Analysis

We obtained 285 nodes and 571 protein pairs from the STRING database by setting combined score to be higher than 0.9 (Fig. **3A**). We identified interaction degrees higher than 15 as hub genes, which were *GNG11* (degree = 19), *HGF* (degree = 17), *SERPINA1* (degree = 17), *GGH* (degree = 17), *SDC2* (degree = 17), *ORM1* (degree = 16) and *CKAP4* (degree = 16).

Afterward, we used Cytohubba to identify hub genes, and the results are shown in Fig. (**3B**). The most significant hub genes revealed by Cytohubba were *SERPINA1, CKAP4, SDC2, SPP1, SERPIND1, CP, PRSS23, SPARCL1, IGFBP3* and *GNG11*.

In addition, MCODE was also utilized for screening hub genes. The most significant modules are shown in Fig. (**3C** and **D**). In Fig. (**3C**), 16 nodes and 65 edges were identified with an MCODE score of 8.677, and in Fig. (**3D**), 17 nodes

and 67 edges were identified with an MCODE score of 9.375 (Fig. **3**).

3.4. Identifying Key Genes for Metastasis

The Venn diagram in Fig. (**3E**) shows four key genes among "significant genes identified by PPI degree", "hub genes identified by Cytohubba" and "hub genes identified by MCODE", including *CKAP4*, *SERPINA1*, *SDC2* and *GNG11*. A summary of these four key genes is shown in Table **2**. The details of the Venn diagram are shown in Table **3**. The gene expression of these four key genes is shown in (Fig. **4**). The results showed that *CKAP4* and *SERPINA1* were downregulated in metastatic cells, while *SDC2* and *GNG11* were upregulated in metastatic cells (Fig. **4B-E**).
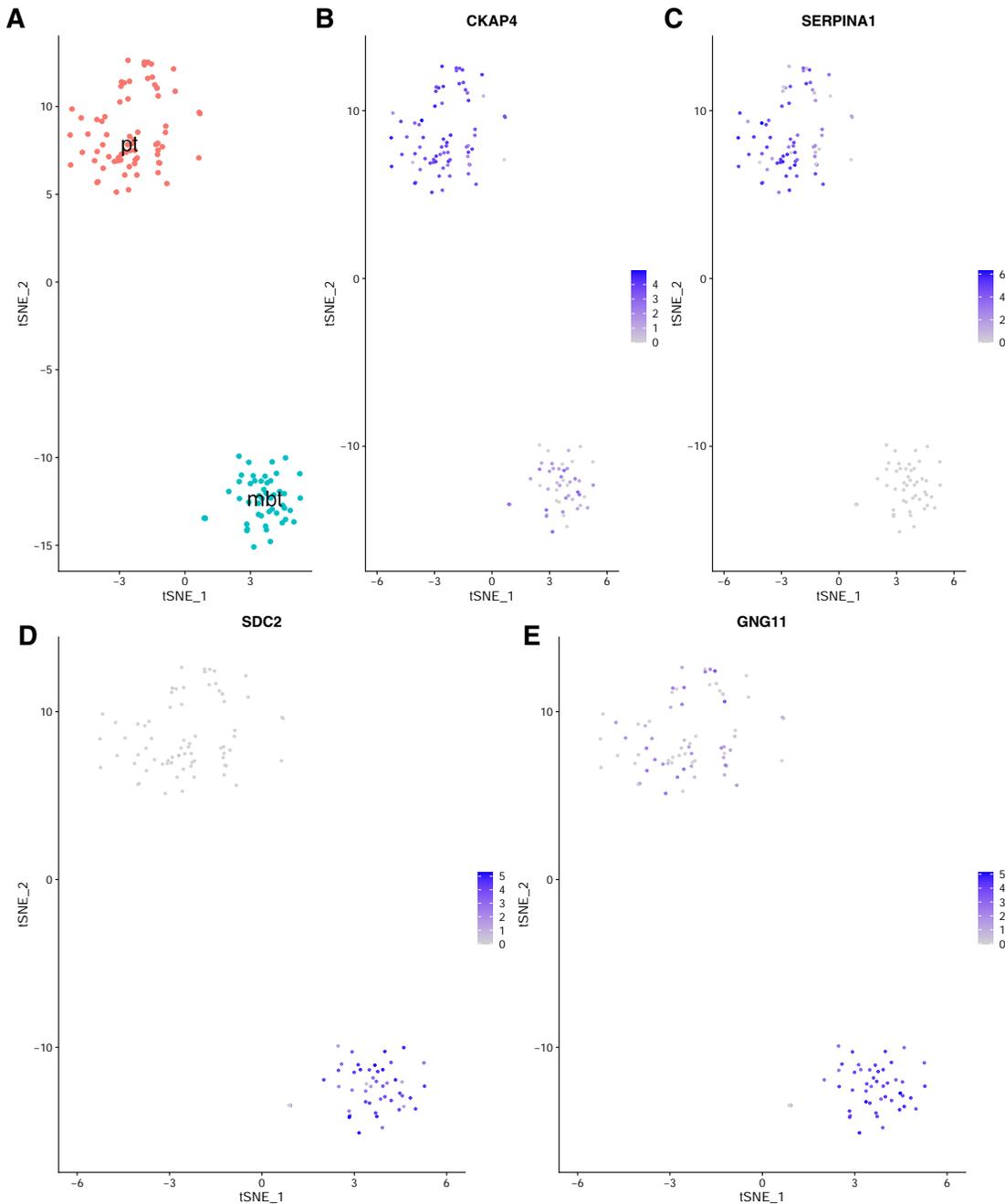
Fig. (4). Gene expression of four key genes identified in this study. (**A**) The nonlinear dimension reduction method t-distributed stochastic neighbour embedding (t-SNE) was applied to scRNA-seq data. Red dots represent primary cells, and blue dots represent metastatic cells. (**B**) Gene expression of key gene *CKAP4*. (**C**) Gene expression of key gene *SERPINA1*. (**D**) Gene expression of key gene *SDC2*. (**E**) Gene expression of key gene *GNG11*. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table 2. Summary of four key genes.

Gene Symbols	Full Names	Functions
CKAP4	Cytoskeleton Associated Protein 4	Mediates the anchoring of the endoplasmic reticulum to microtubules. High-affinity epithelial cell surface receptor for the *FZD8*-related low molecular weight sialoglycopeptide APF/antiproliferative factor. Mediates the APF antiproliferative signalling within cells.
SERPINA1	Serpin Family A Member 1	Inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin and thrombin. Irreversibly inhibits trypsin, chymotrypsin and plasminogen activator. The aberrant form inhibits insulin-induced NO synthesis in platelets, decreases coagulation time and has proteolytic activity against insulin and plasmin.
SDC2	Syndecan 2	Cell surface proteoglycan that bears heparan sulfate. Regulates dendritic arbor morphogenesis.
GNG11	G Protein Subunit Gamma 11	Guanine nucleotide-binding proteins (G proteins) are involved as a modulator or transducer in various transmembrane signalling systems. The beta and gamma chains are required for the GTPase activity, for replacement of GDP by GTP, and for G protein-effector interaction.

Table 3. The VENN diagram results.

Names	Total	Elements
Cytohubba, MCODE, Degree	4	*GNG11*, *CKAP4*, *SERPINA1*, *SDC2*
MCODE, Degree	2	*HGF*, *ORM1*
Cytohubba, MCODE	6	*SPARCL1*, *CP*, *SERPIND1*, *SPP1*, *IGFBP3*, *PRSS23*
Degree	1	*GGH*
MCODE	21	*SERPINE1*, *F8*, *MC4R*, *AGT*, *RAMP1*, *CYSLTR2*, *F2RL1*, *CNR1*, *EDN1*, *ADORA2B*, *SRGN*, *ADRB2*, *GNAI1*, *LPAR2*, *A2M*, PTHLH, *GAL*, *SAA1*, *PCYOX1L*, *GNAT3*, *VIPR1*

4. DISCUSSION

Lung cancer is the leading cause of cancer deaths, and lung adenocarcinoma is the main type of lung cancer (approximately 50%-55%). As the aetiology and pathogenesis of brain metastasis of lung adenocarcinoma are unclear, we designed this study to reveal the key biomarkers of this process. The results of this work present new insight into the biomarkers of brain metastasis of lung adenocarcinoma.

In the present study, single-cell RNA-seq data were utilized to identify DEGs between primary lung tumours and brain metastasis of lung adenocarcinoma. Then, GO and KEGG enrichment analyses were conducted to reveal functional biological pathways related to brain metastasis. Subsequently, four key genes, *CKAP4*, *SERPINA1*, *SDC2* and *GNG11*, were identified by Venn diagram by selecting the common genes among "significant genes identified by PPI degree", "hub genes identified by CytoHubba" and "hub genes identified by MCODE".

CKAP4 (Cytoskeleton Associated Protein 4) is a protein-coding gene that mediates the anchoring of the endoplasmic reticulum to microtubules and is a high-affinity epithelial cell surface receptor for the *FZD8*-related low molecular weight sialoglycopeptide APF/antiproliferative factor. It mediates APF antiproliferative signalling within cells. Yanagita et at. identified *CKAP4* as a novel early serodiagnostic marker for lung cancer [54]. Bhavanasi *et al.* identified *CKAP4* as a receptor for Dickkopf that suppressed tumorigenesis in cancer cells [55]. Li *et al.* found that *CKAP4* inhibited the metastasis of hepatocellular carcinoma by suppressing the activation of epithelial growth factor receptor (EGFR) signalling [56]. Li *et al.* pointed out that *CKAP4* may serve as a

key biomarker of intrahepatic cholangiocellular carcinoma and is significantly associated with distant metastasis [57]. From the above results, the function of *CKAP4* in lung cancer cells is complex, and many studies have shown metastasis in other types of cancers. The function of *CKAP4* in the metastasis of lung cancer requires further study.

SERPINA1 (Serpin Family A Member 1) is a protein-coding gene that serves as an inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin and thrombin. *SERPINA1* irreversibly inhibits trypsin, chymotrypsin and plasminogen activator. The aberrant form inhibits insulin-induced NO synthesis in platelets, decreases coagulation time and has proteolytic activity against insulin and plasmin. Ercetin *et al.* found that the *SERPINA1* gene plays a significant role in the pathogenesis of lung cancer by influencing cancer cell migration and colony formation [58]. Kwon *et al.* pointed out that *SERPINA1* was correlated with lymph node metastasis in colorectal cancer and promoted cancer progression *via* fibronectin [59]. Ortega *et al.* demonstrated the effects of *SERPINA1* on lung function and emphysema using deep gene resequencing [60]. Normandin *et al.* found that *SERPINA1* played a key role in the progression from a primary tumour to invasive metastasis as a protease inhibitor in epithelial ovarian cancer [61]. The findings show that the effect of *SERPINA1* in metastasis is undeniable, but the role of *SEROPINA1* in metastasis in lung adenocarcinoma merits further study.

SDC2 (Syndecan 2) is a protein-coding gene for cell surface proteoglycan that bears heparan sulfate. It regulates dendritic arbor morphogenesis. Huang *et al.* pointed out the relationship between *SDC2* and *CYR61* in regulating the trans-

forming growth factor-beta (TGF-beta) signalling pathway, which plays a significant role in tumour development [62]. Hua *et al.* found that *SDC2* played a carcinogenic role in colorectal cancer by promoting epithelial-mesenchymal transition (EMT) in colorectal cancer cells [63]. Sun *et al.* found that *RKIP* and *HMGA2* regulated the metastasis of breast cancer through lysyl oxidase and *SDC2* [64]. Tsoyi *et al.* demonstrated that *SDC2* silencing *in vivo* reduced lung adenocarcinoma tumour metastasis [65]. Previous studies have already shown important roles of *SDC2* in lung adenocarcinoma tumour metastasis; however, clear mechanisms still need to be determined.

GNG11 (G Protein Subunit Gamma 11) encodes guanine nucleotide-binding proteins (G proteins), which are involved as modulators or transducers in various transmembrane signaling systems. The beta and gamma chains are required for GTPase activity, for GTP replacement by GTP, and for G protein-effector interactions. Hua *et al.* found that *GNG11* acted as a key gene in lung adenocarcinoma; however, it was not associated with survival [66]. Shi *et al.* used bioinformatics tools to identify *GNG11* as a key gene in female lung cancer patients who never smoked associated with poor overall survival [67]. Studies of the role of *GNG11* in lung adenocarcinoma tumour metastasis are rare, and its expression in metastatic cells was significantly upregulated, indicating that activating *GNG11* may be a potential biomarker for prognosis and treatment of metastasis.

CONCLUSION

In summary, the present study identified *CKAP4*, *SERPINA1*, *SDC2,* and *GNG11* as key genes for brain metastasis in lung adenocarcinoma. The results provide new insights into the development between primary tumours and metastatic tumours of lung adenocarcinoma, and these potential biomarkers may lead to better prognosis and treatment of lung adenocarcinoma. However, further molecular biological experiments are still required to confirm the functions of the identified key genes.

AUTHORS' CONTRIBUTIONS

Quan Zou and Zilong Zhang designed the study. Feifei Cui collected the data. Zilong Zhang, Murong Zhou, Song Wu and Bo Gao analyzed the data. Zilong Zhang contributed to the writing of the paper.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Related Code used in this study is available from https://github.com/ZilongZhang44/LADC.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

REFERENCES

[1] Song Q, Shang J, Yang Z, *et al.* Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. J Transl Med 2019; 17(1): 70.
http://dx.doi.org/10.1186/s12967-019-1824-4 PMID: 30832680

[2] Li H, Wang G, Zhang H, *et al.* Prognostic role of the systemic immune-inflammation index in brain metastases from lung adenocarcinoma with different EGFR mutations. Genes Immun 2019; 20(6): 455-61.
http://dx.doi.org/10.1038/s41435-018-0050-z PMID: 30410015

[3] Ghosh A, Yan H. Stability analysis at key positions of EGFR related to non-small cell lung cancer. Curr Bioinform 2020; 15: 260-7.
http://dx.doi.org/10.2174/1574893614666191212112026

[4] Ilhan-Mutlu A, Osswald M, Liao Y, *et al.* Bevacizumab prevents brain metastases formation in lung adenocarcinoma. Mol Cancer Ther 2016; 15(4): 702-10.
http://dx.doi.org/10.1158/1535-7163.MCT-15-0582 PMID: 26809491

[5] Shih DJH, Nayyar N, Bihun I, *et al.* Genomic characterization of human brain metastases identifies drivers of metastatic lung adenocarcinoma. Nat Genet 2020; 52(4): 371-7.
http://dx.doi.org/10.1038/s41588-020-0592-7 PMID: 32203465

[6] Wei C, Dong X, Lu H, *et al.* LPCAT1 promotes brain metastasis of lung adenocarcinoma by up-regulating PI3K/AKT/MYC pathway. J Exp Clin Cancer Res 2019; 38(1): 95.
http://dx.doi.org/10.1186/s13046-019-1092-4 PMID: 30791942

[7] Sun G, Ding X, Bi N, *et al.* Molecular predictors of brain metastasis-related microRNAs in lung adenocarcinoma. PLoS Genet 2019; 15(2): e1007888.
http://dx.doi.org/10.1371/journal.pgen.1007888 PMID: 30707694

[8] Hoj JP, Mayro B, Pendergast AMA. A TAZ-AXL-ABL2 feed-forward signaling axis promotes lung adenocarcinoma brain metastasis. Cell Rep 2019; 29(11): 3421-3434.e8.
http://dx.doi.org/10.1016/j.celrep.2019.11.018 PMID: 31825826

[9] Téglási V, Reiniger L, Fábián K, *et al.* Evaluating the significance of density, localization, and PD-1/PD-L1 immunopositivity of mononuclear cells in the clinical course of lung adenocarcinoma patients with brain metastasis. Neuro Oncol 2017; 19(8): 1058-67.
http://dx.doi.org/10.1093/neuonc/now309 PMID: 28201746

[10] Pocha K, Mock A, Rapp C, *et al.* Surfactant expression defines an inflamed subtype of lung adenocarcinoma brain metastases that correlates with prolonged survival. Clin Cancer Res 2020; 26(9): 2231-43.
http://dx.doi.org/10.1158/1078-0432.CCR-19-2184 PMID: 31953311

[11] Bahrami E, Taheri M, Benam M. Calcified brain metastatic adenocarcinoma: A case report and review of the literature. Neuroradiol J 2019; 32(1): 57-61.
http://dx.doi.org/10.1177/1971400918805184 PMID: 30303450

[12] Shukla S, Evans JR, Malik R, *et al.* Development of a RNA-seq based prognostic signature in lung adenocarcinoma. J Natl Cancer Inst 2016; 109(1): 109.

PMID: 27707839

[13] Yang L, Lv Y, Wang S, *et al.* Identifying FL11 subtype by characterizing tumor immune microenvironment in prostate adenocarcinoma *via* Chou's 5-steps rule. Genomics 2020; 112(2): 1500-15.
http://dx.doi.org/10.1016/j.ygeno.2019.08.021 PMID: 31472243

[14] Wang S, Wang Y, Yu C, *et al.* Characterization of the relationship between FLI1 and immune infiltrate level in tumour immune microenvironment for breast cancer. J Cell Mol Med 2020; 24(10): 5501-14.
http://dx.doi.org/10.1111/jcmm.15205 PMID: 32249526

[15] Peng J, Sun BF, Chen CY, *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res 2019; 29(9): 725-38.
http://dx.doi.org/10.1038/s41422-019-0195-y PMID: 31273297

[16] Park J, Shrestha R, Qiu C, *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science 2018; 360(6390): 758-63.
http://dx.doi.org/10.1126/science.aar2131 PMID: 29622724

[17] Velmeshev D, Schirmer L, Jung D, *et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. Science 2019; 364(6441): 685-9.
http://dx.doi.org/10.1126/science.aav8130 PMID: 31097668

[18] Zhang W, Li W, Zhang J, *et al.* Data integration of hybrid microarray and single cell expression data to enhance gene network inference. Curr Bioinform 2019; 14: 255-68.
http://dx.doi.org/10.2174/1574893614666190104142228

[19] Jiang L, Xiao Y, Ding Y, Tang J, Guo F. FKL-Spa-LapRLS: An accurate method for identifying human microRNA-disease association. BMC Genomics 2018; 19 (Suppl. 10): 911.
http://dx.doi.org/10.1186/s12864-018-5273-x PMID: 30598109

[20] Jiang L, Wang C, Tang J, *et al.* LightCpG: A multi-view CpG sites detection on single-cell whole genome sequence data. BMC Genomics 2019; 20(1): 306.

[21] Liu B, Gao X, Zhang H. BioSeq-Analysis2.0: An updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. Nucleic Acids Res 2019; 47(20): e127.
http://dx.doi.org/10.1093/nar/gkz740 PMID: 31504851

[22] Jin S, Zeng X, Xia F, Huang W, Liu X. Application of deep learning methods in biological networks. Brief Bioinform 2020; 22(2): 1902-17.
http://dx.doi.org/10.1093/bib/bbaa043 PMID: 32363401

[23] Liu X, Hong Z, Liu J, *et al.* Computational methods for identifying the critical nodes in biological networks. Brief Bioinform 2020; 21(2): 486-97.
http://dx.doi.org/10.1093/bib/bbz011 PMID: 30753282

[24] Zeng X, Zhong Y, Lin W, Zou Q. Predicting disease-associated circular RNAs using deep forests combined with positive-unlabeled learning methods. Brief Bioinform 2020; 21(4): 1425-36.
http://dx.doi.org/10.1093/bib/bbz080 PMID: 31612203

[25] Li F, Luo M, Zhou W, *et al.* Single cell RNA and immune repertoire profiling of COVID-19 patients reveal novel neutralizing antibody. Protein Cell 2020.
http://dx.doi.org/10.1007/s13238-020-00807-6 PMID: 33237441

[26] Jiang Q, Wang G, Jin S, Li Y, Wang Y. Predicting human microRNA-disease associations based on support vector machine. Int J Data Min Bioinform 2013; 8(3): 282-93.
http://dx.doi.org/10.1504/IJDMB.2013.056078 PMID: 24417022

[27] Qi R, Ma A, Ma Q, Zou Q. Clustering and classification methods for single-cell RNA-sequencing data. Brief Bioinform 2020; 21(4): 1196-208.
http://dx.doi.org/10.1093/bib/bbz062 PMID: 31271412

[28] Wang Z, Ding H, Zou Q. Identifying cell types to interpret scRNA-seq data: How, why and mre possibilities. Brief Funct Genomics 2020; 19(4): 286-91.
http://dx.doi.org/10.1093/bfgp/elaa003 PMID: 32232401

[29] Kim N, Kim HK, Lee K, *et al.* Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat Commun 2020; 11(1): 2285.
http://dx.doi.org/10.1038/s41467-020-16164-1 PMID: 32385277

[30] Zhang Z, Cui F, Wang C, Zhao L, Zou Q. Goals and approaches for each processing step for single-cell RNA sequencing data. Brief Bioinform 2020; 22(4): bbaa314.

http://dx.doi.org/10.1093/bib/bbaa314 PMID: 33316046

[31] Liang P, Yang W, Chen X, *et al.* Machine learning of single-cell transcriptome highly identifies mRNA signature by comparing F-score selection with DGE analysis. Mol Ther Nucleic Acids 2020; 20: 155-63.
http://dx.doi.org/10.1016/j.omtn.2020.02.004 PMID: 32169803

[32] Cao C, Mak L, Jin G, Gordon P, Ye K, Long Q. PRESM: Personalized reference editor for somatic mutation discovery in cancer genomics. Bioinformatics 2019; 35(9): 1445-52.
http://dx.doi.org/10.1093/bioinformatics/bty812 PMID: 30247633

[33] Li H, Song M, Yang W, Cao P, Zheng L, Zuo Y. A comparative analysis of single-cell transcriptome identifies reprogramming driver factors for efficiency improvement. Mol Ther Nucleic Acids 2020; 19: 1053-64.
http://dx.doi.org/10.1016/j.omtn.2019.12.035 PMID: 32045876

[34] Amezquita RA, Lun ATL, Becht E, *et al.* Orchestrating single-cell analysis with Bioconductor. Nat Methods 2020; 17(2): 137-45.
http://dx.doi.org/10.1038/s41592-019-0654-x PMID: 31792435

[35] Cao C, Kwok D, Edie S, *et al.* kTWAS: Integrating kernel machine with transcriptome-wide association studies improves statistical power and reveals novel genes. Brief Bioinform 2020; bbaa270.
http://dx.doi.org/10.1093/bib/bbaa270 PMID: 33200776

[36] Kim KT, Lee HW, Lee HO, *et al.* Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. Genome Biol 2015; 16: 127.
http://dx.doi.org/10.1186/s13059-015-0692-3 PMID: 26084335

[37] Liu B. BioSeq-Analysis: A platform for DNA, RNA and protein sequence analysis based on machine learning approaches. Brief Bioinform 2019; 20(4): 1280-94.
http://dx.doi.org/10.1093/bib/bbx165 PMID: 29272359

[38] Liu G, Zhang F, Jiang Y, *et al.* Integrating genome-wide association studies and gene expression data highlights dysregulated multiple sclerosis risk pathways. Mult Scler 2017; 23(2): 205-12.
http://dx.doi.org/10.1177/1352458516649038 PMID: 27207450

[39] Jiang Q, Jin S, Jiang Y, *et al.* Alzheimer's disease variants with the genome-wide significance are significantly enriched in immune pathways and active in immune cells. Mol Neurobiol 2017; 54(1): 594-600.
http://dx.doi.org/10.1007/s12035-015-9670-8 PMID: 26746668

[40] Wang G, Wang Y, Feng W, *et al.* Transcription factor and microRNA regulation in androgen-dependent and independent prostate cancer cells. BMC Genomics 2008; 9 (Suppl. 2): S22.
http://dx.doi.org/10.1186/1471-2164-9-S2-S22 PMID: 18831788

[41] Cheng L, Wang P, Tian R, *et al.* LncRNA2Target v2.0: A comprehensive database for target genes of lncRNAs in human and mouse. Nucleic Acids Res 2019; 47(D1): D140-4.
http://dx.doi.org/10.1093/nar/gky1051 PMID: 30380072

[42] Wang G, Wang Y, Teng M, Zhang D, Li L, Liu Y. Signal transducers and activators of transcription-1 (STAT1) regulates microRNA transcription in interferon gamma-stimulated HeLa cells. PLoS One 2010; 5(7): e11794.
http://dx.doi.org/10.1371/journal.pone.0011794 PMID: 20668688

[43] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018; 36(5): 411-20.
http://dx.doi.org/10.1038/nbt.4096 PMID: 29608179

[44] Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: Tool for the unification of biology. Nat Genet 2000; 25(1): 25-9.
http://dx.doi.org/10.1038/75556 PMID: 10802651

[45] Ikram N, Qadir MA, Afzal MT. SimExact - An efficient method to compute function similarity between proteins using gene ontology. Curr Bioinform 2020; 15: 318-27.
http://dx.doi.org/10.2174/1574893614666191017092842

[46] Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 2012; 16(5): 284-7.
http://dx.doi.org/10.1089/omi.2011.0118 PMID: 22455463

[47] Wang S, Zhang Q, Yu C, Cao Y, Zuo Y, Yang L. Immune cell infiltration-based signature for prognosis and immunogenomic analysis in breast cancer. Brief Bioinform 2020; 22(2): 2020-31.
http://dx.doi.org/10.1093/bib/bbaa026 PMID: 32141494

[48] Dennis G Jr, Sherman BT, Hosack DA, *et al.* DAVID: database

for annotation, visualization, and integrated discovery. Genome Biol 2003; 4(5): 3.
http://dx.doi.org/10.1186/gb-2003-4-5-p3 PMID: 12734009

[49]   Szklarczyk D, Morris JH, Cook H, *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017; 45(D1): D362-8.
http://dx.doi.org/10.1093/nar/gkw937 PMID: 27924014

[50]   Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. Genome Biol 2019; 20(1): 185.
http://dx.doi.org/10.1186/s13059-019-1758-4 PMID: 31477170

[51]   Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.
http://dx.doi.org/10.1186/1471-2105-4-2 PMID: 12525261

[52]   Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. CytoHubba: Identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014; 8 (Suppl. 4): S11.
http://dx.doi.org/10.1186/1752-0509-8-S4-S11 PMID: 25521941

[53]   Stelzer G, Rosen N, Plaschkes I, *et al.* The GeneCards suite: From gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016; 54(1): 1.30.1-33.

[54]   Yanagita K, Nagashio R, Jiang SX, *et al.* Cytoskeleton-associated protein 4 is a novel serodiagnostic marker for lung cancer. Am J Pathol 2018; 188(6): 1328-33.
http://dx.doi.org/10.1016/j.ajpath.2018.03.007 PMID: 29751934

[55]   Bhavanasi D, Speer KF, Klein PS. CKAP4 is identified as a receptor for Dickkopf in cancer cells. J Clin Invest 2016; 126(7): 2419-21.
http://dx.doi.org/10.1172/JCI88620 PMID: 27322056

[56]   Li SX, Liu LJ, Dong LW, *et al.* CKAP4 inhibited growth and metastasis of hepatocellular carcinoma through regulating EGFR signaling. Tumour Biol 2014; 35(8): 7999-8005.
http://dx.doi.org/10.1007/s13277-014-2000-3 PMID: 24838946

[57]   Li MH, Dong LW, Li SX, *et al.* Expression of cytoskeleton-associated protein 4 is related to lymphatic metastasis and indicates prognosis of intrahepatic cholangiocarcinoma patients after surgery resection. Cancer Lett 2013; 337(2): 248-53.
http://dx.doi.org/10.1016/j.canlet.2013.05.003 PMID: 23665508

[58]   Ercetin E, Richtmann S, Delgado BM, *et al.* Clinical significance of *SERPINA1* gene and its encoded alpha1-antitrypsin protein in NSCLC. Cancers 2019; 11(9): 11.
http://dx.doi.org/10.3390/cancers11091306 PMID: 31487965

[59]   Kwon CH, Park HJ, Choi JH, *et al.* Snail and serpinA1 promote tumor progression and predict prognosis in colorectal cancer. Oncotarget 2015; 6(24): 20312-26.
http://dx.doi.org/10.18632/oncotarget.3964 PMID: 26015410

[60]   Ortega VE, Li X, O'Neal WK, *et al.* The effects of rare *SERPINA1* variants on lung function and emphysema in SPIROMICS. Am J Respir Crit Care Med 2020; 201(5): 540-54.
http://dx.doi.org/10.1164/rccm.201904-0769OC PMID: 31661293

[61]   Normandin K, Péant B, Le Page C, *et al.* Protease inhibitor SERPINA1 expression in epithelial ovarian cancer. Clin Exp Metastasis 2010; 27(1): 55-69.
http://dx.doi.org/10.1007/s10585-009-9303-6 PMID: 20049513

[62]   Huang X, Xiao DW, Xu LY, *et al.* Prognostic significance of altered expression of SDC2 and CYR61 in esophageal squamous cell carcinoma. Oncol Rep 2009; 21(4): 1123-9.
PMID: 19288017

[63]   Hua R, Yu J, Yan X, *et al.* Syndecan-2 in colorectal cancer plays oncogenic role *via* epithelial-mesenchymal transition and MAPK pathway. Biomed Pharmacother 2020; 121: 109630.
http://dx.doi.org/10.1016/j.biopha.2019.109630 PMID: 31707342

[64]   Sun M, Gomes S, Chen P, *et al.* RKIP and HMGA2 regulate breast tumor survival and metastasis through lysyl oxidase and syndecan-2. Oncogene 2014; 33(27): 3528-37.
http://dx.doi.org/10.1038/onc.2013.328 PMID: 23975428

[65]   Tsoyi K, Osorio JC, Chu SG, *et al.* Lung adenocarcinoma syndecan-2 potentiates cell invasiveness. Am J Respir Cell Mol Biol 2019; 60(6): 659-66.
http://dx.doi.org/10.1165/rcmb.2018-0118OC PMID: 30562054

[66]   Hua P, Zhang Y, Jin C, Zhang G, Wang B. Integration of gene profile to explore the hub genes of lung adenocarcinoma: A quasi-experimental study. Medicine 2020; 99(43): e22727.
http://dx.doi.org/10.1097/MD.0000000000022727   PMID: 33120770

[67]   Shi K, Li N, Yang M, Li W. Identification of key genes and pathways in female lung cancer patients who never smoked by a bioinformatics analysis. J Cancer 2019; 10(1): 51-60.
http://dx.doi.org/10.7150/jca.26908 PMID: 30662525