Identification of core genes and clinical outcomes in tumors originated from endoderm (gastric cancer and lung carcinoma) via bioinformatics analysis

Yewen Shi, MD, PhD\textsuperscript{a}, Dongmin Chang, MD, PhD\textsuperscript{b}, Wenhan Li, MD, PhD\textsuperscript{c,d}, FengYu Zhao, MD\textsuperscript{b}, Xiaoyong Ren, MD, PhD\textsuperscript{a}, Bin Hou, MD, PhD\textsuperscript{d,e,\textsuperscript{*}}

Abstract
During last decade, bioinformatics analysis has provided an effective way to study the relationship between various genes and biological processes. In this study, we aimed to identify potential core candidate genes and underlying mechanisms of progression of lung and gastric carcinomas which both originated from endoderm. The expression profiles, GSE54129 (gastric carcinoma) and GSE27262 (lung carcinoma), were collected from GEO database. One hundred eleven patients with gastric carcinoma and 21 health people were included in this research. Meanwhile, there were 25 lung carcinoma patients. Then, 75 differentially expressed genes were selected via GEO2R online tool and Venn software, including 31 up-regulated genes and 44 down-regulated genes. Next, we used Database for Annotation, Visualization, and Integrated Discovery and Metascape software to analyze Kyoto Encyclopedia of Gene and Genome pathway and gene ontology. Furthermore, Cytoscape software and MCODE App were performed to construct complex of these differentially expressed genes. Twenty core genes were identified, which mainly enriched in extracellular matrix-receptor interaction, focal adhesion, and PI3K-Akt pathway \((P < .01)\). Finally, the significant difference of gene expression between cancer tissues and normal tissues in both lung and gastric carcinomas was examined by Gene Expression Profiling Interactive Analysis database. Twelve candidate genes with positive statistical significance \((P < .01)\), COMP CTHRC1 COL1A1 SPP1 COL11A1 COL10A1 CXCL13 CLDN3 CLDN1 matrix metalloproteinases 7 ADAM12 PLAU, were picked out to further analysis. The Kaplan–Meier plotter website was applied to examine relationship among these genes and clinical outcomes. We found 4 genes (ADAM12, SPP1, COL1A1, COL11A1) were significantly associated with poor prognosis in both lung and gastric carcinoma patients \((P < .05)\). In conclusion, these candidate genes may be potential therapeutic targets for cancer treatment.

Abbreviations: BP = biological process, CC = cellular component, DAVID = Database for Annotation, Visualization, and Integrated Discovery, ECM = extracellular matrix, EMT = epithelial–mesenchymal transition, GC = gastric cancer, GEPIA = Gene Expression Profiling Interactive Analysis, GO = gene ontology, KEGG = Kyoto Encyclopedia of Gene and Genome, MF = molecular function, PPI = protein–protein interaction network, SCLC = small-cell lung cancer, SPP1 = secreted phosphoprotein 1.

Keywords: bioinformatics analysis, differentially expressed genes (DEGs), gastric carcinoma, lung carcinoma

1. Introduction
Cancer is one of the major public health problems and the second leading cause of death in the world.\textsuperscript{[1]} Thus, identifying biomarkers in the development of cancers is a crucial part to facilitate new therapies. As we know, the endoderm is the earliest cell types which provide origination of majority human internal systems, including respiratory and gastrointestinal tubes.\textsuperscript{[2]} Meanwhile, many shared molecular signaling pathways, TGF-
β pathway, Wnt pathway, FGF pathway, and so on, are also involved in the formation and elaboration of internal organs which evolved from endoderm. Besides, lung carcinoma and gastric carcinoma are the 3th and the 7th death reasons respectively in China. Therefore, this research would like to explore whether or not several common genes play pivotal roles in the development of lung and gastric carcinomas, which both originated from endoderm.

During the last decade, with the rapid development in high-throughput technologies, Gene-chip has become an effective method to quickly detect differentially expressed genes (DEGs). Moreover, bioinformatics analysis could provide genetic background, potential molecular mechanisms, relationship between genes and clinical outcomes of patients. Therefore, increasing evidences are explored for new researches and cancer treatment based on these data. Although several studies have identified some core genes in the development of lung cancer or gastric cancer, the common features of lung and gastric cancers that evolved from endoderm cells are not explored.

In the present study, we obtained 2 original microarray datasets, GSE54129 from patients with gastric carcinoma and GSE27262 from patients with lung carcinoma, in GEO database. DEGs were selected via online tool GEO2R. Database for Annotation, Visualization, and Integrated Discovery (DAVID) and Metascape software were performed to analyze these DEGs including cellular component (CC), biological process (BP), molecular function (MF), and underlying signaling pathway. Subsequently, the protein–protein interaction network (PPI) was constructed by STRING database, the hub of genes was identified via Cytoscape software and MCODE App. Twenty candidate genes were verified after further analyzing. Finally, we used the Gene Expression Profiling Interactive Analysis (GEPIA) and Kaplan–Meier plotter tools to examine the relationship among genes expression and clinical outcomes. Surprisingly, we found 4 genes (ADAM12, secreted phosphoprotein 1 [SPP1], encoding Collagen I alpha-1 chain protein [COL1A1], COL1A1) played pivotal roles in the development of both lung and gastric carcinomas. The 4 key DEGs may act as potential biomarkers for early diagnosis or as underlying targets for cancer treatment.

2. Methods

2.1. Ethics

The patients’ clinical prognosis and animal research were not involved in this observation study. We downloaded materials and data from public database which were introduced in following content. Therefore, it is unnecessary for us to state ethics approval.

2.2. Microarray data selection

NCBI-GEO database (www.ncbi.nlm.nih.gov/geo) is a free public resource, including gene expression data of different tumors. To identify whether or not some core genes play an important role in tumors originated from endoderm (gastric carcinoma and lung carcinoma), we obtained 2 expression profiling data sets of primary cancer tissues and normal tissues, GSE54129 (gastric carcinoma) and GSE27262 (lung carcinoma), from GEO database. Microarray data of GSE54129 and GSE27262 were all on account of GPL570 Platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). There were 111 samples from patients with subtotal gastrectomy and 21 samples from health people. Besides, tumor and adjacent normal tissue pairs from 25 lung adenocarcinoma patients were selected in this research.

2.3. Identification of significant genes

GER2R was an online tool and used to analyze DEGs, |logFC| > 1.5 and adjust P value < 0.05 were regarded as cutoff criteria. Then, the raw data from 2 microarrays were collected for further analysis. TXT format data were processed in the Venn diagrams website (http://bioinformatics.psb.ugent.be/venn/10) to select common genes among the 2 data sets. The genes with logFC > 0 were regarded as up-regulated genes, while the value of down-regulated genes was logFC < 0.

2.4. GO analysis and KEGG pathway analysis

Gene ontology (GO) ontology analysis is widely used in bioinformatics, which contains 3 parts of biological features, BP, CC, and MF. Kyoto Encyclopedia of Gene and Genome (KEGG) enrichment is also performed to identify potential biological pathways of DEGs. Thus, we used public databases DAVID (https://david.ncifcrf.gov/) and Metascape (http://metascape.org/gp/index.html#main/step1) which provided online tools for genes annotation, visualization, and integrated discovery function to analyze DEGs. The cutoff value was adjust P < 0.05.

2.5. PPI network and module analysis

The STRING database (Search Tool for the Retrieval of Interacting Genes https://string-db.org) is commonly performed to predict and trace out PPI network in bioinformatics. We uploaded significant genes into website and downloaded information for further analysis. Cytoscape (version 3.6.0), a free visualization software, was used to examine relationship between these DEGs (maximum number of interactors = 0 and confidence score ≥0.4). Meanwhile, MCODE software in Cytoscape was applied to screen central nodes in PPI network (degree cut-off = 2, node score cut-off = 0.2, k-core = 2, and maximum depth = 100).

2.6. Expression levels of hub genes and clinical outcomes analysis

GEPIA (http://gepia.cancer-pku.cn/) is an online tool based on TCGA and GTEx database, was used to identify mRNA expression level of DEGs. The genes whose expression data had statistical difference in both 2 tumors were selected to further analyze. Kaplan–Meier-plotter (http://www.kmplot.com) is a user-friendly, interactive web resource to predict prognosis of patients with different cancer types (breast cancer, gastric cancer, lung cancer, and ovarian cancer). Thus, we applied Kaplan–Meier-plotter website to estimate the effect of candidate genes on patients’ clinical outcomes based on GEO, EGA, and TCGA databases. The hazard ratio with 95% confidence intervals and log rank P value were calculated and showed on the plot.

3. Results

3.1. Identification of core genes in tumors originated from endoderm

Our research included 111 gastric carcinoma samples, 21 health gastric mucosa samples, 25 lung carcinoma tissues, and 25 adjacent normal tissues. After analyzing microarrays via GEO2R
online tool, 1672 different genes and 1053 different genes were extracted from GSE54129 and GSE27262 respectively. Then 75 common DEGs were selected via Venn Diagram Software, including 31 up-regulated (logFC > 0) genes and 44 down-regulated (logFC < 0) genes (as shown in Fig. 1 and Table 1).

3.2. DEGs ontology and KEGG pathway analysis in tumors originated from endoderm

All common DEGs were analyzed using DAVID and Metascape online tools. As shown in Table 2, the results of GO BP analysis in DAVID software demonstrated that 31 up-regulated genes mainly enriched in collagen fibril organization, cell adhesion, proximal/distal pattern formation, embryonic limb morphogenesis, and cellular response to amino acid stimulus. Meanwhile, 44 down-regulated genes mainly enriched in BMP signaling pathway, negative regulation of cortisol biosynthetic process, negative regulation of cell proliferation, negative regulation of insulin-like growth factor receptor signaling pathway and growth. For the GO CC group, up-regulated genes were enriched in collagen trimer, extracellular space, and proteinaceous extracellular matrix; Then different genes in down-regulated group mainly enriched in extracellular space and exosome; In the BP group, 31 up-regulated genes particularly enriched in heparin binding, identical protein binding, platelet-derived growth factor binding, and calcium ion binding. Besides, down-regulated genes significantly enriched in BMP receptor binding and transforming growth factor beta receptor binding. Interestingly, the results of GO ontology from Metascape website provided another variation in the BP of DEGs (Fig. 2). Seventy-five core genes from GSE54129 and GSE27262 positively enriched in sketal system development, cellular response to growth factor stimulus, embryonic morphogenesis, bone remodeling, and blood vessel development. KEGG analysis results of DEGs were exhibited in Table 3 which indicated that core genes were particularly enriched in extracellular matrix (ECM)–receptor interaction, focal adhesion, Amoebiasis, protein digestion and absorption, PI3K-Akt signaling pathway, and platelet activation (P < .01).

3.3. Identification of core candidate genes with PPI (protein–protein interaction) network and modular analysis

A total of 75 DEGs, including 31 up-regulated genes and 44 down-regulated genes, were imported into STRING website and Cytoscape software. The results that contained 41 nodes and 78 edges were further analyzed by MCODE app. Finally, 20 core candidate genes were identified among 41 genes, including 16 up-regulated genes and 4 down-regulated genes (as shown in Fig. 3).

3.4. Reanalysis of 20 selected genes by KEGG pathway enrichment

To identify potential pathway of 20 selected genes, KEGG pathway enrichment was reanalyzed via DAVID database. Results demonstrated that 6 genes (COMP, COL1A2, COL1A1, COL11A1, THBS2, SPP1) mainly enriched in ECM–receptor interaction, focal adhesion and PI3K-Akt pathway (P < .01, Table 4).

3.5. Analysis of core genes via the GEPIA and Kaplan–Meier plotter

GEPIA online tool was performed to analyze the expression level of 20 core genes. The results demonstrated that there were 12

---

**Table 1**

All 75 differentially expressed genes (DEGs) were selected from 2 GSE datasets, including 31 up-regulated genes and 44 down-regulated genes.

| DEGs        | Genes name                                                                 |
|-------------|-----------------------------------------------------------------------------|
| Up-regulated| ADAM12, PLAU, MMP7, IGFBP3, ANP3E, CXCL13, COL1A1, COL19, PDGFD, CST1, HDX10, SPP1, GREM1, BUB1, COL3A1, CLDN3, HDX10, COMP, THBS2, WISP1, F12, FDCSP, SULF1, DOK5, ENC1, CLDN1, COL10A1, COL11A1, CTB, COL1A2, PTPR8, MRRF, DOR |
| Down-regulated| OTUD1, FRMD3, GATA6, MAL, AHNAK, KL4, TMEM246, ABCA5, ADK4, NEDD4L, ASPA, BMP2, ADRB2, FKB21, KANK4, LAMA3, MR68372, SEMA4B, ADGR6, NCKAP5, PLOC4, MYRF, CLIC3, ADMA1TS, ADGR13, CA2, ADTRP, SSTR1, MYZAP, PLLP, HIP, PSC, SOSTDC1, LRPR, RMST, PIP5K1B, SNN1H, KL11, MANC2C, BMPS5, CLDN18, ABHD6, GNB2, HVL1 |
core genes with significant difference among 20 candidates (as shown in Table 5, Figs. 4 and 5). The genes whose expression had statistical significance in both gastric and lung carcinomas were imported into KM-plotter website. Then, the prognostic information of 12 different genes was available in KM-plotter website. As exhibited in Table 6 and Figures 6 and 7, 4 genes (ADAM12, SPP1, COL1A1, COL11A1) were all associated with a poor prognosis of gastric cancer and lung cancer (P < 0.05).

4. Discussion

According to the 2018 Global Cancer Statistics,[14] lung carcinoma is the most commonly diagnosed cancer (11.6% of the total cases) and the leading cause of cancer death (18.4% of the total cancer deaths), and gastric carcinoma is the 5th cancer for incidence (5.7%) and 3rd for mortality (8.2%) in worldwide. Meanwhile, the 3 leading incident cancers and the leading cause of cancer death among both men and women are lung, stomach, and liver cancers in China.[6]

Recently, many studies have demonstrated the similarity and relationship between embryo development and tumorigenesis with regard to invasive cellular behaviors, gene expression, and other important biological behaviors. Meanwhile, cells driving from endoderm show a surprising number of shared features and basic biological processes, including epithelial–mesenchymal transition (EMT), cell migration.[11,13] As we know, lung and stomach both originated from endoderm during embryonic development. To explore common features in the development of lung and gastric carcinomas, we present a comprehensive bioinformatics analysis. Finally, we have identified 4 key genes (ADAM12, SPP1, COL1A1, COL11A1) that were positively associated with the clinical poor prognosis.

ADAM-12, also known as meltrin-α, is a disintegrin and metalloproteinase family member that plays important roles in embryonic development. ADAM-12 is capable of doing several biological functions, including proteolysis, regulation of growth factor availability, cell-cell and cell-matrix adhesion, cell signaling, and ectodomain shedding.[11,12] In the majority of normal tissues, the expression of ADAM12 is extremely low.[12,13] But its expression will increase in certain pathological conditions, leguchi et al[14] identified that ADAM12 enhanced ephrin-A1 cleavage which deteriorated the EphA1/ephrin-A1-mediated cell adhesion in the lung cancer with endocrine manner. Finally, this biological process resulted in causing lung hyper-permeability and contributed to cancer metastasis.[14] A report showed ADAM12 was highly expressed in small cell lung cancer (SCLC) and was an effective biomarker for cancer diagnosis and prognostic prediction.[15,16] Moreover, E2F1 transcription factor regulated the expression of ADAM12 by binding differential cis-acting elements.[17] Carl-McGrath et al[18] found ADAM9, 12 were over-expressed in gastric cancer in the malignant growth of gastric cancer cells via the interaction with adhesion molecules or the proteolytic “sheding” of signaling molecules.

SPP1 is a secreted glyco-phosphoprotein which is involved in the attachment of osteoclasts to mineralize bone matrix. This protein is also a cytokine that up-regulates expression of interferon-gamma and interleukin-12.[15] Over-expression of SPP1 was highly associated with aggressive phenotypes and TNM stages of lung cancer and promotes SCLC proliferation viability, inhibits apoptosis and autophagy.[20,21] Besides, Wang et al[22] suggested that SPP1 enhanced the second-generation EGFR TKI resistance in lung cancer. Thus, inhibiting the expression of SPP1 might be an effective therapy to overcome afatinib resistance. Previous data have highlighted the critical role of SPP1 in gastric cancer (GC),[23] SPP1 was a target gene of miR-340 which could mediate the PI3K/AKT signaling pathway in GC.[24]

COL1A1 (encoding Collagen I alpha-1 chain protein), belonging to the collagen family, is a major structural component of the ECM and epithelial tumorigenesis. Overexpression of COL1A1 is significantly correlated with maintaining lung cancer progression. To further explore the correlation between the expression of core genes with significant difference among 20 candidates (as shown in Table 5, Figs. 4 and 5).
cell growth in 3D culture. Oleksiewicz et al. found that COL1A1 was extremely overexpressed in non-small cell lung cancer (NSCLC) compared with adjacent normal tissues. Therefore, it was regarded as a novel player in the complex network of hypoxia response in NSCLC. While, Wang and Yu proved that miR-129-5p could suppress gastric cancer cell

| Pathway ID | Name                                | Count | %    | P value | Genes                                      |
|------------|-------------------------------------|-------|------|---------|--------------------------------------------|
| hsa04512   | ECM-receptor interaction             | 8     | 0.1  | 5.7E-8  | LAMA3, COMP, COL3A1, COL1A2, COL1A1, THBS2, COL11A1, SPP1 |
| hsa04510   | Focal adhesion                       | 8     | 0.1  | 2.0E-5  | LAMA3, COMP, COL3A1, COL1A2, COL1A1, THBS2, COL11A1, SPP1 |
| hsa05146   | Amoebiasis                           | 6     | 0.1  | 7.6E-5  | PLOB4, LAMA3, COL3A1, COL1A2, COL1A1, COL11A1 |
| hsa04974   | Protein digestion and absorption     | 5     | 0.1  | 4.2E-4  | COL3A1, COL1A2, COL1A1, COL11A1, COL10A1 |
| hsa04151   | PI3K-Akt signaling pathway           | 8     | 0.1  | 4.4E-4  | LAMA3, COMP, COL3A1, COL1A2, COL1A1, THBS2, COL11A1, SPP1 |
| hsa04611   | Platelet activation                  | 5     | 0.1  | 2.0E-3  | PLOB4, COL3A1, COL1A2, COL1A1, COL11A1 |

Table 3

KEGG pathway analysis of different genes.

KEGG = Kyoto Encyclopedia of Gene and Genome.
proliferation, migration, and invasion through inhibiting expression of COL1A1. Interestingly, COL1A1 mRNA expression was positively up-regulated in premalignant and malignant tissues in gastric cancer and highly associated with poor prognosis.[28] Meanwhile, COL11A1, another member of the collagen family encoding collagen XI alpha-1 chain protein, is significantly correlated with the pathological stage and metastasis of NSCLC.[29] COL11A1 promotes cell proliferation, migration, invasion, and drug resistance mediated by Smad signaling in recurrent NSCLC.[30] The same biological functions of COL11A1 could be observed in gastric cancer.[31,32] In our study, COL1A1 and COL11A1 are enriched in ECM-receptor interaction pathways.[33] Main constituents of ECM include collagens, proteoglycans, and noncollagenous glycoproteins. ECM not only plays a modulatory role in epigenetic alteration and specification during embryonic development,[34] but also could be a part of tumor-micro-environment to support tumor growth. Amedei et al.[35] suggest that HP0157 provides a compact link between \textit{H. pylori} and gastric cancer via matrix degradation. Meanwhile, collagen I, as an important element of ECM, induces EMT via TGF-\beta signaling in NSCLC cell lines,[36] which is known to play an important role in cancer progression, metastasis, and drug resistance.

**Table 5**

| Category | Genes |
|----------|-------|
| Genes with statistical difference in gastric cancers ($P < .01$) | ADAM12 CLDN1 CLDN3 COL1A1 COL1A2 COL10A1 COL11A1 COMP CTHRC1 CXCL13 MMP7 PLAU SPP1 THBS1 WISP1 |
| Genes with statistical difference in lung cancers ($P > .01$) | ACKR4 ADAM12 BMP2 CCL19 CLDN1 CLDN3 CLDN18 COL1A1 COL10A1 COL11A1 COMP CTHRC1 CXCL13 MMP7 PLAU SPP1 SSTR1 |
| Genes with statistical difference in 2 cancers ($P < .01$) | COMP CTHRC1 COL1A1 SPP1 COL1A1 COL10A1 CXCL13 CLDN3 CLDN1 MMP7 ADAM12 PLAU |
| Genes without statistical difference in 2 cancers ($P > .01$) | COL1A2 BMP2 CCL19 ACKR4 SSTR1 CLDN18 THBS2 WISP1 |
resistance. Moreover, it has been proved that skeletal development and cell adhesion module increases the invasion of cancer cell by COL11A1 activation to SPP1.[37] In general, COL1A1, COL11A1, and SPP1, as core genes in this study, suggest that they play a critical role in ECM–receptor interaction and cell adhesion signaling pathway and may form a network in lung and stomach cancers.

In recent years, the target therapy in cancer has attracted many researchers’ attentions. This new treatment uses monoclonal antibodies and small-molecule drugs to target specific
genes and proteins that are involved in the growth of cancer cells. While targeted therapies are specific to a type of cancer and works better for some types of cancer than others.[38,39] Considering the similarity and relationship between embryo development and tumorigenesis, we tried to use a comprehensive bioinformatics analysis to find some core genes and common features in the development of lung cancer and gastric cancer, which could provide valuable materials to develop novel target therapies for lung and gastric cancer treatment.

Figure 5. The expression levels of 20 core candidate genes in lung cancer, ** means cut-off \( P < .01 \).
Table 6

The prognostic information of the 12 key genes in both 2 cancers.

| Category                                                                 | Genes                                                                 |
|--------------------------------------------------------------------------|----------------------------------------------------------------------|
| Genes with significantly worse survival in gastric cancer ($P < .05$)   | ADAM12, CLDN3, COL1A1, COL10A1, COL11A1, COMP, CTHRC1, MMP7, PLAU, SPP1 |
| Genes with significantly worse survival in lung cancer ($P > .05$)       | ADAM12, COL1A1, COL11A1, SPP1                                         |
| Genes with significantly worse survival in both 2 cancer types ($P < .05$) | ADAM12, COL1A1, SPP1, COL11A1                                       |
| Genes without significantly worse survival ($P > .05$)                   | COMP, COL10A1, CTHRC1, CXCL13, CLDN3, CLDN1, MMP7, PLAU              |

Figure 6. The clinical outcomes of 12 core genes in patients with gastric cancer. $P$ value was shown in each figure. The cut-off value was $P < .05$. 
Finally, there are some limitations in our research. One of the major weaknesses is limited samples in our study. The reason of this problem is that we have selected several candidates from GEO database, such as GSE63571, GSE31552, GSE29249 (lung cancer expression profiles) and GSE38932, GSE27342, GSE20143 (gastric cancer expression profiles) etc., but various difficulties hampered our further analysis. For example, different platforms were used in different database; therefore, some gene symbols cannot be used as unified criterion, GSE31552 was analyzed in platform GPL6244, and GPL5936 platform was

Figure 7. The prognostic information of 12 core genes in patients with lung cancer, P value was shown in each figure. The cut-off value was P < .05.
applied for GSE38932 analysis. Thus, we collected GSE54129 (gastric carcinoma) and GSE27262 (lung carcinoma) data which were from same platform with unified criterion. Besides, some expression profiles contained few cases, like GSE63371 just had 3 primary cancer samples and 2 metastatic cancer samples. Furthermore, a lot of data cannot be downloaded for further analysis. Second, more bioinformatics methods and database should be used in this research including R language, Ocombe, COSMIC database, which will influence outcomes of this research. Therefore, it is necessary to learn more knowledge about bioinformatics. Then, as mentioned above, many other tumors are derived from organs of endodermal origin. Those cancer types should be taken into account. For exploring potential mechanisms, our team planned to use other tumor types from endoderm in future researches. Besides, the next step of our study is to do experiments in vivo and vitro to confirm the results in the present study. It will be helpful to uncover underlying mechanisms of cancer progression from specific germ layer.

5. Conclusion
We identified 4 DEGs (ADAM12, SPPI, COL1A1, COL1A1A) via a comprehensive bioinformatics analysis, which may be involved in the progress and clinical outcomes of patients with lung carcinoma and gastric carcinoma. We concluded that ECM-receptor interaction and cell adhesion signaling pathway may affect progresses and clinical outcomes of tumors (lung carcinoma and gastric carcinoma) originated from endoderm. Our investigation could provide valuable information about oncogenesis and progression. And result can give some new inspiration for target therapy in different cancers. Further experimental studies are also needed to confirm these results which could help determine potential therapeutic targets of tumors from same germ layer in embryonic development.

Author contributions
Conceptualization: Yewen Shi, Bin Hou.
Data curation: Yewen Shi, Wenhua Li.
Formal analysis: Wenhua Li, Fengyu Zhao.
Funding acquisition: Xiaoyong Ren.
Methodology: Wenhua Li, Fengyu Zhao, Bin Hou.
Project administration: Bin Hou.
Resources: Dongmin Chang, Xiaoyong Ren.
Software: Fengyu Zhao.
Supervision: Dongmin Chang, Xiaoyong Ren, Bin Hou.
Writing – original draft: Yewen Shi, Bin Hou.
Writing – review & editing: Yewen Shi, Dongmin Chang, Bin Hou.

References
[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
[2] Nowotschin S, Hadjantonakis AK, Campbell K. The endoderm: a divergent cell lineage with many commonalities. Development 2019;146:dev150920.
[3] Ferry JK, Lins RJ, Lobie PE, et al. Regulation of invasive growth: similar epigenetic mechanisms underpin tumour progression and implantation in human pregnancy. Clin Sci (Lond) 2009;118:451–7.
[4] Patra SK, Deb M, Patra A. Molecular marks for epigenetic identification of developmental and cancer stem cells. Clin Epigenetics 2011;2:27–53.
[5] Riddoli L, Petrimi M, Flammingh I, et al. Human embryo immune escape mechanisms rediscovered by the tumor. Immunobiology 2009;214:61–76.
[6] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–32.
[7] Bremner C. Applications of bioinformatics in cancer. Cancers (Basel) 2019;11:1630.
[8] Chen J, Coppola G. Bioinformatics and genomic databases. Handb Clin Neurol 2018;147:75–92.
[9] Singer J, Irmisch A, Ruscheweyh HJ, et al. Bioinformatics for precision oncology. Brief Bioinform 2019;20:778–88.
[10] Can T. Introduction to bioinformatics. Methods Mol Biol 2014;1107:51–71.
[11] Nieto MA. Epithelial plasticity: a common theme in embryonic and cancer cells. Science 2013;342:1234830.
[12] Wheelock MJ, Johnson KR. Cadherins as modulators of cellular phenotype. Annu Rev Cell Dev Biol 2003;19:207–35.
[13] Jacobsen J, Visse R, Sorensen HP, et al. Catalytic properties of ADAM12 and its domain deletion mutants. Biochemistry 2008;47:537–47.
[14] Ieguchi K, Tomita T, Omori T, et al. ADAM-12 cleaved ephrin-A1 contributes to lung metastasis. Oncogene 2004;23:2179–90.
[15] Shao S, Li Z, Gao W, et al. ADAM-12 as a diagnostic marker for the proliferation, migration and invasion in patients with small cell lung cancer. PLoS One 2014;9:e85396.
[16] Mino N, Miyahara R, Nakayama F, et al. A disintegrin and metalloprotease 12 (ADAM12) is a prognostic factor in resected pathological stage I lung adenocarcinoma. J Surg Oncol 2009;100:267–72.
[17] Li Z, Wang Y, Kong L, et al. Expression of ADAM12 is regulated by E2F1 in small cell lung cancer. Oncol Rep 2015;34:3231–7.
[18] Carl-Mcgrath S, Lendeckel U, Ebert M, et al. The disintegrin-metallproteinases ADAM9, ADAM12, and ADAM15 are upregulated in gastric cancer. Int J Oncol 2005;26:17–24.
[19] Wu Q, Zhang B, Sun Y, et al. Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by integrated microarray analysis. Onco Targets Ther 2019;12:1545–63.
[20] Hu Z, Lin D, Yuan J, et al. Overexpression of osteopontin is associated with more aggressive phenotypes in human non-small cell lung cancer. Clin Cancer Res 2005;11:4646–52.
[21] Liu H, Wei S, Zhang L, et al. Secretedp1 promotes the development of small cell lung cancer cells by inhibiting autophagy and apoptosis. Pathol Oncol Res 2019;25:1487–95.
[22] Wang X, Zhang F, Yang X, et al. Secreted phosphoprotein 1 (SPPI) contributes to second-generation EGFR tyrosine kinase inhibitor resistance in non-small cell lung cancer. Oncol Res 2019;27:971–7.
[23] Chen LZ, He CY, Su X, et al. SPPI rs5734 and its epistatic interactions with SPARC polymorphisms in gastric cancer susceptibility. Gene 2018;640:43–50.
[24] Song SZ, Lin S, Liu JN, et al. Targeting of SPPI by macroRNA-340 inhibits gastric cancer cell epithelial-mesenchymal transition through inhibition of the PI3K/AKT signaling pathway. J Cell Physiol 2019;234:18587–601.
[25] Li J, Li X, Lan T, et al. [Type I collagen secreted by lung cancer cells promotes cancer cell growth in a three-dimensional culture system]. Nan Fang Yi Ke Da Xue Xue Bao 2014;34:1129–34.
[26] Oleksiewicz U, Liloglou T, Tasopoulou KM, et al. COL1A1, PRPF40A, and UCP2 correlate with hypoxia markers in non-small cell lung cancer. J Cancer Res Clin Oncol 2017;143:1133–41.
[27] Wang Q, Yu J. MiR-129-5p suppresses gastric cancer cell invasion and proliferation by inhibiting COL1A1. Biochem Cell Biol 2018;96:19–35.
[28] Chen J, Coppola G. Bioinformatics and genomic databases. Handb Clin Neurol 2018;147:75–92.
[29] Li A, Li J, Lin J, et al. COL11A1 is overexpressed in gastric cancer tissues and its domain deletion mutants. Biochemistry 2008;47:537–47.
[30] Ieguchi K, Tomita T, Omori T, et al. ADAM-12 cleaved ephrin-A1 contributes to lung metastasis. Oncogene 2004;23:2179–90.
[31] Shi et al. Medicine (2021) 100:12 www.md-journal.com
[34] Malta D, Reticker-Flynn NE, Da SC, et al. Extracellular matrix microarrays to study inductive signaling for endoderm specification. Acta Biomater 2016;34:30-40.

[35] Amedei A, Munari F, Bella CD, et al. Helicobacter pylori secreted peptidyl prolyl cis, trans-isomerase drives Th17 inflammation in gastric adenocarcinoma. Intern Emerg Med 2014;9:303-9.

[36] Shintani Y, Maeda M, Chaika N, et al. Collagen I promotes epithelial-to-mesenchymal transition in lung cancer cells via transforming growth factor-beta signaling. Am J Respir Cell Mol Biol 2008;38:95-104.

[37] Sun Y, Wang L, Jiang M, et al. Secreted phosphoprotein 1 upstream invasive network construction and analysis of lung adenocarcinoma compared with human normal adjacent tissues by integrative bio computation. Cell Biochem Biophys 2010;56:59-71.

[38] Marme D. Advances in cancer therapy: targeted therapies. Oncol Res Treat 2016;39:758-9.

[39] Gasser M, Waaga-Gasser AM. Therapeutic antibodies in cancer therapy. Adv Exp Med Biol 2016;917:95-120.