Identification of Hub Genes, Modules and Metabolic Pathways Associated With Lung Adenocarcinoma: A System Biology Approach

Raheleh Roudi

Department of Cell Systems and Anatomy, University of Texas Health San Antonio, San Antonio, TX, USA

Behnaz Beikzadeh

Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

Giandomenico Roviello

Department of Health Sciences, University of Florence, Florence, Italy

Alberto D'angelo

Department of Biology and Biochemistry, University of Bath, Bath, B2 7AY, UK

Morteza Hadizadeh ([✉ morteza.hadizade@gmail.com](mailto:morteza.hadizade@gmail.com))

Physiology Research Center, Institute of Basic and Clinical Physiology Sciences, Kerman University of Medical Sciences, Kerman, Iran

Research Article

Keywords: lung cancer, adenocarcinoma, diagnostic markers, prognostic factors, meta-analysis

Posted Date: November 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-111252/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Lung cancer is the most common and fatal malignant tumour worldwide with a five-year overall survival rate of only 15%. Lung adenocarcinoma (LUAD) is a heterogeneous disease. The use of microarray datasets along with bioinformatics knowledge might help to clarify the expression profile of cancer, molecular markers for the initial screening of tumour and the underlying biological mechanisms. The present study is designed to identify differential expression genes and molecular mechanisms of LUAD compared to normal lung tissues using systems biology approaches.

Methods Four GSE datasets (GSE75037, GSE63459, GSE32863 and GSE10072) were selected from the Gene Expression Omnibus (GEO) database. Data processing and meta-analysis were performed using the R statistical programming language. The differentially expressed genes (DEGs) associated with each stage were obtained. The common and unique DEGs between stages of LUAD and adjacent normal lung tissues were initiated by Venny tool. Common genes, including upregulated and downregulated genes, were then analyzed to a STRING database to obtain protein-protein interaction (PPI). STRING output was analyzed by MCODE and CytoHubba applications of Cytoscape to identify modules of co-expression and hub genes, respectively.

Results The shared upregulated and downregulated DEGs among LUAD stages were 22 and 140 genes, respectively, when compared to normal lung tissues. Unique genes for each stage were also identified. The hub genes were PECAM1, TEK, CDH5, VWF and ANGPT1. Most of the top cluster genes were enriched for Gq(s) signalling events, GPCR ligand binding, class B/2 (Secretin family receptors), platelet activation, signalling and aggregation in the main three co-expression clusters. Most of the shared genes (16 genes) were enriched in the metabolic pathway of hemostasis. Meaningful signaling pathways for unique genes were found at each stage.

Conclusions In the present study main three co-expression clusters, metabolic pathways and biological processes were identified to understand mechanisms underlying LUAD pathogenesis, development and progression at different stages. Unique upregulated and downregulated DEGs at each stage were identified with FERMT1 and SIX1 as specific early-stage diagnostic biomarkers for stage IB and IIB. 5 hub genes were observed, including PECAM1, TEK, CDH5, VWF and ANGPT1 which might be crucial for the onset and progression of LUAD.

Introduction

Lung cancer remains the most prevalent and fatal malignant tumour on a global scale with a five-year overall survival rate of only 15%¹. Nearly 85% of lung cancer cases are categorized into non-small cell lung cancer (NSCLC) which includes adenocarcinoma and squamous cell carcinoma^{2,3}. Lung adenocarcinoma (LUAD) is the most common histological NSCLC subtype with an estimated frequency of approximately 40 percent⁴. During the last decade, early-stage detection of LUAD using low-dose computed tomography (LDCT) and treatment modalities including surgical intervention followed by chemo/radiotherapy regimens have been used; notwithstanding, recurrence is a common event, ranging from 30 to 75 percent. In addition, more than 80 percent of relapses occur within 2 years after conventional treatments^{5,6}. Surprisingly, significant improvements have been achieved in LUAD treatment as reported by molecular targeted therapy and immunotherapy such as the immune checkpoint inhibitors. However, the incidence rate of LUAD is showing an alarming trend within non-smokers at younger ages^{7,8}. The insight that LUAD grows slower than other histological subtypes - as well as with higher probability of detection prior to - has sparked significant interest among scientists⁴.

LUAD is a heterogeneous disease¹¹ and the examination of microarray-based expression profiles could identify thousands of genes independently, while the use of network analysis based on system biology can define transcription networks and key genes involved in disease¹². The use of microarray datasets along with bioinformatics knowledge might help to clarify the expression profile of cancers and, consequently, different biological mechanisms can be obtained¹³. Several studies have shown that molecular markers can be used for the initial screening of tumours¹⁴. Candidate upregulated and downregulated genes selected by bioinformatics analysis can be assayed in biochemical analysis and subsequent steps¹⁵.

Growing evidence has shown that previous studies were conducted to determine the genetic changes in the patients with LUAD^{16,17}. However, these studies examined only a limited number of patients and restricted genetic dysregulations or geographical variation¹⁸. Therefore, the application of computational systems biology approach might help to shed lighting on the molecular aspects of LUAD. Discovering the molecular mechanisms involved in LUAD can help to identify the pathogenesis of this cancer. On the other hand, finding specific LUAD biomarkers in any stage of the disease can help with the screening and early diagnosis. The present study is designed to identify differentially expressed genes and molecular mechanisms of LUAD compared to normal lung tissues using systems biology approaches. Lastly, gene expression patterns at each stage of LUAD, ranging from I to IV, plus common and definite dysregulated genes related to each stage were represented.

Methods

1) Eligibility criteria and dataset selection

The gene datasets were screened by two independent authors (R. R. and M. H.) and those deemed relevant were selected according to the inclusion and exclusion criteria. The datasets were identified according to the following inclusion criteria: differential gene expression analysis of primary LUAD compared to normal lung tissues without any prior treatment and data availability on the definite (certain) stage of patients. The gene datasets with specific categorization were excluded.

2) Selection of gene expression datasets for meta-analysis

The mRNA expression datasets of LUAD were searched using the keywords: 'lung cancer' and 'Homo sapiens'[porgn: txid9606], and 'Expression profiling by array' against the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>). After a systematic review, four GSE datasets (GSE75037,

GSE63459, GSE32863 and GSE10072) were selected and downloaded using the GEO2R database. While GSE75037 and GSE32863 were based on GPL6884 (Illumina HumanWG-6 v3.0 expression beadchip), GSE63459 and GSE10072 were based on GPL6883 (Illumina HumanRef-8 v3.0 expression beadchip) and GPL96 (HG-U133A) Affymetrix Human Genome U133A Array), respectively (Table 1). Three GSE profiles, including GSE31267 (categorization of patients based on the smoking status), GSE31210 (categorization based on ALK, EGFR, and KRAS mutations) and GSE20189 (analysis on blood samples) were excluded. All data were publically available, and the present study did not involve any human or animal experimentation.

2) Processing of GEO dataset and identification of differentially expressed genes

GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) is a web-based instrument on GEOquery and limma packages from the Bioconductor project, which can compare two or more groups of specimens to recognize differentially expressed genes (DEGs). In the current study, each dataset was assigned either LUAD or normal tissue and GEO2R was then applied to identify DEGs between them. The occurrence of false-positive results was corrected using adjusted p values (adj. p value (Benjamini-Hochberg procedure)). The adj. p value < 0.05 and $\log_2FC > |1|$ were set as cut-off values.

3) Microarray data processing and integrative meta-analysis

data processing and integration procedures were performed using the R statistical programming language. The aforementioned four datasets made use of different and very popular platforms (Illumina and Affymetrix). A sensitive step in the integration of heterogeneous data is normalization (2). Before data merging, each dataset was first normalized using the Limma package. The Surrogate Variable Analysis (SVA) package was used for the removal of batch effects (non-biological differences) using the ComBat function (3). Batch effect removal was checked by Principal Component Analysis (PCA) and boxplot. The outcome of the meta-analysis is a unit expression matrix, namely a combination of four datasets of this study based on common Entrez IDs. We extrapolated DEGs associated with each stage from the unit expression matrix. For the definition of DEGs, $\log_2$ fold change $\geq |1.5|$ and adjusted p-values ≤ 0.001 were used. Given that the data were acquired from multiple microarray platforms, different pre-processing approaches were applied for each platform.

4) Co-expression network construction in LUAD and adjacent normal lung tissues

The co-expression network and hub genes at different tumour stages of LUAD and adjacent normal lung tissues are initially processed by determining the common DEGs by Venny tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). The list of common DEGs was then subjected to a STRING database to obtain protein-protein interaction (PPI). The outcome of STRING was analyzed by MCODE and CytoHubba (using the Maximal Clique Centrality (MCC) function) plugin of Cytoscape to identify modules of co-expression and hub genes respectively. Lastly, top modules and hub genes were selected according to the highest score.

Results

Dataset selection, differential gene expression analysis and identification of DEGs

Four gene expression microarray datasets, including GSE75037, GSE63459, GSE32863 and GSE10072 were selected for our study. The raw data were obtained from 219 normal lung adjacent tissue and 232 LUAD patients with a wide range of tumour stages, from IA and IB to IV (Table 1). The GSE75037 dataset consisted of 83 LUAD patients versus 83 controls, while the dataset GSE63459 consisted of 33 LUAD patients versus 32 controls. GSE32863 and GSE10072 datasets consisted of 58 LUAD patients versus 58 controls, and 58 LUAD patients versus 49 controls, respectively. Before integrative meta-analysis, each dataset was analyzed and genes with $\log_2$ fold change $\geq |1|$ and adjusted p-value threshold of 0.05 were considered as significantly differentially expressed.

After dataset merging, one expression matrix with 451 samples was obtained, including 219 controls, 62 samples at stage IA, 68 cases at stage IB. Also, we found 21 and 38 specimens at stages IIA and IIB, respectively. Thirty-three samples were stage IIIA, whereas only 4 cases were observed at stage IIIB and 6 samples at stage IV (Table 1). DEGs of each stage were obtained from the unit expression matrix using $\log_2$ fold change $\geq |1.5|$ and adjusted p-values ≤ 0.001 . Comparison of significant differences between control and stage IA: 314 genes (75 upregulated genes, 239 downregulated genes); control versus stage IB: 341 genes (89 upregulated genes, 252 downregulated genes); control versus stage IIA: 437 genes (130 upregulated genes, 307 downregulated genes); control versus stage IIB: 385 genes (108 upregulated genes, 277 downregulated genes); control versus stage IIIA: 434 genes (119 upregulated genes, 315 downregulated genes); control versus stage IIIB: 307 genes (94 upregulated genes, 213 downregulated genes); control versus stage IV: 533 genes (206 upregulated genes, 327 downregulated genes) were obtained.

Venny Analysis

After independent analysis of significant upregulated and downregulated DEGs at each stage with Venny software, the shared upregulated and downregulated DEGs among different stages of LUAD cases compared to normal lung tissues were 22 and 140, respectively (Table 2). The expression level of shared 162 genes among all stages of LUAD cases relative to normal lung tissues is shown in figure 1.

Unique genes for each stage were also identified, with specific details reported in Table 3. Our findings support the idea that the highest number of unique DEGs genes between is reported by LUAD patients with stage IV, including 84 upregulated and 55 downregulated genes. Detailed information on venny analysis can be found in Supplementary File 1.

PPI Network, Module and hub genes Analysis

At this stage, shared genes (162 common DEGs) were used to elaborate a protein network (protein-protein interaction) with the STRING database (<https://string-db.org/>). Subsequently, using the Molecular Complex Detection (MCODE) and CytoHubba applications, co-expression clusters and the highly

connected hub genes were identified respectively (Fig 2). Table 4 shows the top three co-expression clusters. The hub genes included PECAM1, TEK, CDH5, VWF and ANGPT1. Gene enrichment analysis of the top three cluster genes was performed by ToppGene platform. Most of the top cluster genes were reported within metabolic pathways such as Gα (s) signalling events, GPCR ligand binding, class B/2 (Secretin family receptors), Platelet activation, signalling and aggregation (Table 5).

GO and pathway enrichment analysis

In addition to enrichment analysis, genes that were common to all stages were further analyzed using ToppGene tool. Most of the genes (16 genes in total) were found to be enriched in the metabolic pathway of Hemostasis and other superior metabolic pathways including Extracellular matrix organization, Platelet degranulation, Response to elevated platelet cytosolic Ca²⁺, Cell surface interactions at the vascular wall and Gα (s) signaling events. Information of the aforementioned metabolic pathways are reported in Table 6.

Biological processes from gene ontology (GO) analysis were also examined with the ToppGene site. The top 25 biological processes can be seen in Table 7. Most genes are involved in the regulation of cell population proliferation whereas cell adhesion, biological adhesion, and response to endogenous stimulus were the processes that involved the most genes. Details of biological processes are available in supplementary file 2.

Pathway enrichment analysis for unique genes

As previously mentioned, Table 3 reports, top Gene was utilized to conduct Reactome pathways enrichment analysis of unique DEGs. In stage IA for the CRLF1 gene, IL-6-type cytokine receptor ligand interactions (FDR = 0.02) and Interleukin-6 family signaling (FDR = 0.02) pathways were significant. There were important pathways for the MFAP5 gene, elastic-associated molecules (FDR = 0.02), elastic fibre formation (FDR = 0.0) and finally for the AKR1B10 gene, metabolic pathways for retinal metabolism and transport (FDR = 0.02), fat-soluble vitamin metabolism (FDR = 0.02) and visual phototransduction (FDR = 0.03). In Stage IB, there was only the FERMT1 gene for which no meaningful pathway was found. The signaling pathways associated with the unique genes in Stage IIA were Translocation of ZAP-70 to Immunological synapse (FDR = 0.04), Phosphorylation of CD3 and TCR zeta chains (FDR = 0.04) and PD-1 signaling (FDR = 0.04). In Stage IIB, there was only the SIX1 gene, which did not find a significant path based on the Reactome database, but from the KEGG database, the pathway of transcriptional misregulation in cancer (FDR = 0.01) was found. The only significant signaling pathway from the unique genes in Stage IIIA was *Listeria monocytogenes* entry into host cells (FDR = 0.04). No significant path was found in Stage IIIB. Finally, in Stage IV, there was only the signaling pathway of serotonin metabolism (FDR = 0.04).

Discussion

Lung cancer is the most common malignant tumour, with the highest cancer-related mortality rate (22%)¹⁹ and a five-year overall survival rate of approximately 15%¹. For these reasons, lung cancer has become of major interest in clinical research²⁰. The two main histotypes of lung cancer include non-small cell lung cancer (NSCLC) (85%) and small cell lung cancer (SCLC) (15%). NSCLC is further classified into three categories: adenocarcinoma, squamous cell carcinoma and large cell with adenocarcinoma the most common subtype of NSCLC²¹. Due to the poor prognosis of NSCLC, investigation of molecular mechanisms responsible for NSCLC development together with the discovery of biomarkers are of great importance for early detection²². One Promising strategy for the identification of main biomarkers is to investigate DEGs in a disease.

In the present study, four microarray datasets including GSE10072, GSE32863, GSE63459 and GSE75037 were analyzed. After merging the four mentioned datasets, DEGs at each LUAD stage, in comparison with the control (normal lung tissues), were identified from a single expression matrix. The shared upregulated and downregulated DEGs among all LUAD stages were 22 and 140 genes, respectively. Gene set enrichment analysis for shared genes showed that most of the genes (16 genes) were within the metabolic pathway of Hemostasis, the other superior metabolic pathways were: Extracellular matrix organization, Platelet degranulation, Response to elevated platelet cytosolic Ca²⁺, Cell surface interactions at the vascular wall and Gα (s) signaling events. Tumor cells or tumor-related inflammatory cells can stimulate the homeostasis system through expression of pre-coagulation proteins, exposure to pre-coagulant lipids, release of inflammatory cytokines, and adhesion to host vascular cells. Coagulation activation and tumor progression are closely linked (Coagulation and cancer: biological and clinical aspects/Analysis of haemostasis biomarkers in patients with advanced stage lung cancer during hypofractionated radiotherapy treatment). Unique upregulated and down-regulated DEGs for each stage were also identified. Comparison of stage IB and IIB showed only one upregulated gene for each stage, which included FERMT1 and SIX1, respectively. FERMT1 and SIX1 can be examined as specific biomarkers for stage IB and IIB.

Fermitin family member 1 (FERMT1, Kindlin-1) is a focal adhesion protein and an epithelial-specific regulator of integrin functions²³ and reported to be expressed in 60% of lung cancers. Although overexpressed in the NSCLC epithelium, FERMT1 is correlated with the differentiation of lung cancer²⁴. In lung cancer, Kindlin-1 downregulates Axin2, a component of the wnt signaling pathway, and upregulates Claudin-1 and -3 (tight junction molecules)²⁵. SIX1 is a homeodomain transcription factor and a downstream target of Notch2²⁶. In one study, Xia Y et al. reported that SIX1 increased NSCLC invasion and proliferation. Upregulation of SIX1 was observed at both mRNA and protein levels by Mima T et al., making way for lung adenocarcinoma invasion²⁷. These two biomarkers can be used to identify the preliminary stages of the disease.

Meaningful signaling pathways for unique upregulated and downregulated DEGs (table 3) were examined using the ToppGene database and the Reactome database at each stage. In stage IA for the CRLF1 gene, IL-6-type cytokine receptor ligand interactions and Interleukin-6 family signaling pathways, for the MFAP5 gene, the pathways of Molecules associated with elastic, Elastic fibre formation and finally for the AKR1B10 gene, metabolic pathways of Retinoid metabolism and transport, Metabolism of fat-soluble vitamins and Visual phototransduction were significant. The signaling pathways associated with the unique genes in Stage IIA were Translocation of ZAP-70 to Immunological synapse, Phosphorylation of CD3 and TCR zeta chains and PD-1 signaling. In Stage

IIB, there was only the SIX1 gene; the pathway of transcriptional misregulation in cancer was found from the KEGG database. The only significant signaling pathway from the unique genes in Stage IIIA was *Listeria monocytogenes* entry into host cells. Finally, in Stage IV, there was only the signaling pathway of serotonin metabolism.

Among the shared genes of all stages, a protein network was plotted, subsequently top 3 co-expression clusters and 5 hub genes were identified with the MCODE and CytoHubba plugin, respectively. Most of the top cluster genes were enriched in metabolic pathways such as Gα(s) signaling events, GPCR ligand binding, class B/2 (Secretin family receptors), Platelet activation, signaling and aggregation. The hub genes included PECAM1, TEK, CDH5, VWF and ANGPT1 of which the CDH5 and VWF genes were in the third co-expression cluster. These 5 hub genes were part of the downregulated DEGs for all stages.

PECAM-1 (platelet endothelial cell adhesion molecule-1) is a 130 kDa transmembrane glycoprotein also known as CD31²⁸. This protein is a member of the immunoglobulin superfamily and is commonly expressed on the surface of platelets, monocytes, neutrophils, endothelial cells, and a large population of circulating T lymphocytes^{28,29}. Functions of PECAM-1 include leukocyte migration, angiogenesis, integrin activation and the intercellular junction of endothelial cells³⁰. Although phosphorylation of PECAM-1 is associated with aggregate-dependent growth and cell adhesion, its degradation reduces tumour cell size and the colony-forming ability³¹. PECAM1 is involved in LUAD tumorigenesis by regulating the expression of vascular endothelial growth factor. In fact, it can maintain and restore vascular integrity and is commonly used as a marker for angiogenesis^{28,32}.

The TIE receptor family, which incorporates TIE1 and TIE2 and is likewise referred to as TEK, plays an essential function in vasculature formation³³ and is mainly expressed in endothelial cells³⁴. In the study of Tao Peng et al., TEK has been reported to be downregulated in LUAD tissues and cell lines, indicating poor patient survival; additionally, overexpression of TEK in LUAD cell lines showed impaired proliferation and aggression abilities^{35,36}. TEK was identified in 2018 as the gene involved in angiogenesis, cell growth, cytokine secretion and inflammatory response in the lung adenocarcinoma³⁶.

CDH5 or VE-cadherin is a transmembrane protein that is responsible for cell–cell adhesion of Endothelial cells (ECs)³⁷ and promotes the integration of ECs³⁸. In lung cancer, the expression of VEGF and CDH5 in circulating microvesicles is correlated with distant metastasis. Researchers have found that CDH5 expression in NSCLC tissues is related to node metastasis and poor prognosis³⁹. CDH5 is important for angiogenesis in tumour and promotes tumour progression via the transforming growth factor – β signalling pathway⁴⁰.

Von Willebrand factor (vWF) is a large multi-subunit glycoprotein that leads to platelet adhesion and accumulation to the sub-endothelium. vWF is observed in platelets, endothelial cells and megakaryocytes⁴¹. This molecule plays a pivotal role in homeostasis. Not only it helps platelets to be absorbed in damaged endothelium but also latest research displays its tumour angiogenesis, inflammation and metastasis potential.^{42,43} On the surface of tumour cells, several vWF receptors such as glycoprotein Ib, integrin αIIbβ3 and αVβ3 have been identified; however, a direct link between vWF and tumour cells has also been observed⁴⁴. vWF is a highly specific marker for vascular endothelial cells⁴⁵. In many cancers, high levels of vWF have been reported in plasma and tumour microenvironment, indicating an advanced stage of the disease. vWF facilitates tumour cell growth by upregulating inflammatory signals, angiogenesis and vascular permeability⁴³.

Angiopoietins belong to a growth factor family required for the formation of blood vessels, including ANG-1, which is the Tie2 Receptor ligand⁴⁶. ANG-1 binding of Tie2 induces Tie2 autophosphorylation, which leads to vascular integration⁴⁶. The overall effect of ANG-1, ANG-2 and TIE-2 on tumour include angiogenesis, inflammation and vascular injection⁴⁷. ANGPT1 maintains normal distribution of PECAM1 and junctional adhesion molecule A (JAMA) in tumour vessel walls⁴⁸.

The top 25 biological processes from gene ontology (GO) analysis showed that most genes were involved in the regulation of cell population proliferation; cell adhesion, biological adhesion, and response to endogenous stimulus were the processes in which most genes were subsequently involved.

The results of this study, obtained through bioinformatics analysis based on the system biology approach, provide details on potential diagnostic biomarkers, genes and signalling pathways involved in LUAD pathogenesis. In order to verify these results, further studies and molecular biological experiments are needed.

Conclusions

22 upregulated and 140 downregulated shared DEGs, based on the analysis of 4 microarray datasets, were identified among all stages of LUAD when compared to normal lung tissues. Unique upregulated and downregulated DEGs for each stage were also identified. FERMT1 and SIX1 can be used as a specific biomarker for stage IB and IIB to diagnose early-stage disease. 5 hub genes were obtained including PECAM1, TEK, CDH5, vWF and ANGPT1 which might be crucial for the onset and progression of LUAD. Top co-expression clusters, metabolic pathways and biological processes were identified which might help to better understand the underlying mechanism in LUAD development and progression at different stages.

All methods were carried out in accordance with relevant guidelines and regulations.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors read and approved the final manuscript.

Availability of data and materials

The cDNA microarray data are publicly available in Gene Expression Omnibus, accession number (GSE75037, GSE63459, GSE32863 and GSE10072).

Competing interests

The authors declare that they have no conflict of interest.

Funding

No.

Authors' contributions

RR and MH conceptualized and designed the study.

RR and MH collected the data.

RR and MH screened and extracted the data. RR and MH conducted data analysis.

All authors reviewed the data analysis.

RR, BB and MH drafted the article.

RR, GR and AD critically revised the article.

All authors have read and approved the final article.

Acknowledgements

Not applicable.

I confirm that I understand BMC Bioinformatics is an open access journal that levies an article processing charge per articles accepted for publication. By submitting my article I agree to pay this charge in full if my article is accepted for publication.

No, I declare that the authors have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration (from you or one of your Contributing Authors) by another publisher.

I have read the BMC journal policies on author responsibilities and submit this manuscript in accordance with those policies.

All of the material is owned by the authors and/or no permissions are required.

I am the author responsible for the submission of this article and I accept the conditions of submission and the BMC Copyright and License Agreement as detailed above.

References

1. Zhu, J. *et al.* Immunotherapy (excluding checkpoint inhibitors) for stage I to III non-small cell lung cancer treated with surgery or radiotherapy with curative intent. *Cochrane Database of Systematic Reviews* (2017) doi:10.1002/14651858.cd011300.pub2.
2. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
3. Network, T. C. G. A. R. & The Cancer Genome Atlas Research Network. Erratum: Corrigendum: Comprehensive molecular profiling of lung adenocarcinoma. *Nature* vol. 514 262–262 (2014).
4. Zappa, C. & Mousa, S. A. Non-small cell lung cancer: current treatment and future advances. *Translational Lung Cancer Research* vol. 5 288–300 (2016).
5. Williams, B. A. *et al.* Predicting postrecurrence survival among completely resected nonsmall-cell lung cancer patients. *Ann. Thorac. Surg.* **81**, 1021–1027 (2006).
6. Sugimura, H. *et al.* Survival After Recurrent Nonsmall-Cell Lung Cancer After Complete Pulmonary Resection. *The Annals of Thoracic Surgery* vol. 83 409–418 (2007).
7. Takeuchi, K. *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nature Medicine* vol. 18 378–381 (2012).
8. Hashimoto, T. *et al.* Different subtypes of human lung adenocarcinoma caused by different etiological factors. Evidence from p53 mutational spectra. *Am. J. Pathol.* **157**, 2133–2141 (2000).

9. Wen, P. *et al.* Identification of candidate biomarkers and pathways associated with SCLC by bioinformatics analysis. *Mol. Med. Rep.***18**, 1538–1550 (2018).
10. Nakamura, H. *et al.* Identification of key modules and hub genes for small-cell lung carcinoma and large-cell neuroendocrine lung carcinoma by weighted gene co-expression network analysis of clinical tissue-proteomes. *PLOS ONE* vol. 14 e0217105 (2019).
11. Selvaraj, G. *et al.* Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. *Journal of Biomedical Informatics* vol. 86 120–134 (2018).
12. Li, Y. *et al.* Network-based approach identified cell cycle genes as predictor of overall survival in lung adenocarcinoma patients. *Lung Cancer***80**, 91–98 (2013).
13. Virtanen, C. & Woodgett, J. Clinical uses of microarrays in cancer research. *Methods Mol. Med.***141**, 87–113 (2008).
14. Qiu, J., Sun, M., Wang, Y. & Chen, B. Identification of Hub Genes and Pathways in Gastric Adenocarcinoma Based on Bioinformatics Analysis. *Med. Sci. Monit.***26**, e920261 (2020).
15. Kihara, D., Yang, Y. D. & Hawkins, T. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. *Cancer Inform.***2**, 25–35 (2007).
16. Zhang, C. *et al.* Genomic Landscape and Immune Microenvironment Features of Preinvasive and Early Invasive Lung Adenocarcinoma. *J. Thorac. Oncol.***14**, 1912–1923 (2019).
17. Sui, J. *et al.* Molecular characterization of lung adenocarcinoma: A potential four-long noncoding RNA prognostic signature. *J. Cell. Biochem.***120**, 705–714 (2019).
18. Jiang, L. *et al.* Association between the novel classification of lung adenocarcinoma subtypes and EGFR/KRAS mutation status: A systematic literature review and pooled-data analysis. *European Journal of Surgical Oncology* vol. 45 870–876 (2019).
19. Jones, G. S. & Baldwin, D. R. Recent advances in the management of lung cancer. *Clin. Med.* **18**, s41–s46 (2018).
20. Cheng, B. *et al.* An annual review of the remarkable advances in lung cancer clinical research in 2019. *J. Thorac. Dis.***12**, 1056–1069 (2020).
21. Duma, N., Santana-Davila, R. & Molina, J. R. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin. Proc.***94**, 1623–1640 (2019).
22. Piao, J. *et al.* Target gene screening and evaluation of prognostic values in non-small cell lung cancers by bioinformatics analysis. *Gene***647**, 306–311 (2018).
23. Mahawithitwong, P. *et al.* Kindlin-1 expression is involved in migration and invasion of pancreatic cancer. *Int. J. Oncol.***42**, 1360–1366 (2013).
24. Zhan, J. *et al.* Opposite role of Kindlin-1 and Kindlin-2 in lung cancers. *PLoS One***7**, e50313 (2012).
25. Sin, S. *et al.* Role of the focal adhesion protein kindlin-1 in breast cancer growth and lung metastasis. *J. Natl. Cancer Inst.***103**, 1323–1337 (2011).
26. Mimae, T. *et al.* Upregulation of notch2 and six1 is associated with progression of early-stage lung adenocarcinoma and a more aggressive phenotype at advanced stages. *Clin. Cancer Res.***18**, 945–955 (2012).
27. Liu, Q. *et al.* The expression profile and clinic significance of the SIX family in non-small cell lung cancer. *J. Hematol. Oncol.***9**, 119 (2016).
28. Kuang, B.-H. *et al.* The prognostic value of platelet endothelial cell adhesion molecule-1 in non-small-cell lung cancer patients. *Med. Oncol.***30**, 536 (2013).
29. Zheng, L.-F., Chen, P., Wang, Z., Wang, X.-J. & Lei, L. Preliminary study on the role of serum PECAM-1 in metastatic breast cancer. *Transl. Cancer Res.***6**, 1207–1213 (2017).
30. Tang, D. G. *et al.* Identification of PECAM-1 in solid tumour cells and its potential involvement in tumour cell adhesion to endothelium. *J. Biol. Chem.***268**, 22883–22894 (1993).
31. Zhang, X., Xu, L.-H. & Yu, Q. Cell aggregation induces phosphorylation of PECAM-1 and Pyk2 and promotes tumour cell anchorage-independent growth. *Mol. Cancer***9**, 7 (2010).
32. Yu, D.-H. *et al.* Effects of hub genes on the clinicopathological and prognostic features of lung adenocarcinoma. *Oncol. Lett.***19**, 1203–1214 (2020).
33. Davis, S. *et al.* Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning. *Cell***87**, 1161–1169 (1996).
34. Kontos, C. D., Cha, E. H., York, J. D. & Peters, K. G. The endothelial receptor tyrosine kinase Tie1 activates phosphatidylinositol 3-kinase and Akt to inhibit apoptosis. *Mol. Cell. Biol.***22**, 1704–1713 (2002).
35. Gál, Z. *et al.* Investigation of the Possible Role of Tie2 Pathway and TEK Gene in Asthma and Allergic Conjunctivitis. *Frontiers in Genetics* vol. 11 (2020).
36. Peng, T., Yang, F., Sun, Z. & Yan, J. TEK Suppresses Lung Adenocarcinoma Cell Phenotypes by Interacting with miR-19a-3p. (2020).
37. Yamaoka-Tojo, M. *et al.* IQGAP1 mediates VE-cadherin-based cell–cell contacts and VEGF signaling at adherence junctions linked to angiogenesis. *Arterioscler. Thromb. Vasc. Biol.***26**, 1991–1997 (2006).
38. Vestweber, D. VE-cadherin: the major endothelial adhesion molecule controlling cellular junctions and blood vessel formation. *Arterioscler. Thromb. Vasc. Biol.***28**, 223–232 (2008).
39. Hung, M.-S. *et al.* Epidermal Growth Factor Receptor Mutation Enhances Expression of Cadherin-5 in Lung Cancer Cells. *PLoS One***11**, e0158395 (2016).
40. Mao, X.-G. *et al.* CDH5 is specifically activated in glioblastoma stemlike cells and contributes to vasculogenic mimicry induced by hypoxia. *Neuro. Oncol.***15**, 865–879 (2013).
41. Schwachtgen, J. L. *et al.* Ets transcription factors bind and transactivate the core promoter of the von Willebrand factor gene. *Oncogene***15**, 3091–3102 (1997).
42. Xu, Y. *et al.* GATA3-induced vWF upregulation in the lung adenocarcinoma vasculature. *Oncotarget***8**, 110517–110529 (2017).

43. Patmore, S., Dhami, S. P. S. & O'Sullivan, J. M. Von Willebrand factor and cancer; metastasis and coagulopathies. *J. Thromb. Haemost.* (2020) doi:10.1111/jth.14976.

44. Guo, R., Yang, J., Liu, X., Wu, J. & Chen, Y. Increased von Willebrand factor over decreased ADAMTS-13 activity is associated with poor prognosis in patients with advanced non-small-cell lung cancer. *J. Clin. Lab. Anal.***32**, e22219 (2018).

45. Regina, S. *et al.* Tissue factor expression in non-small cell lung cancer: relationship with vascular endothelial growth factor expression, microvascular density, and K-ras mutation. *J. Thorac. Oncol.***3**, 689–697 (2008).

46. Thomas, M. & Augustin, H. G. The role of the Angiopoietins in vascular morphogenesis. *Angiogenesis***12**, 125–137 (2009).

47. Shim, W. S. N., Ho, I. A. W. & Wong, P. E. H. Angiopoietin: A TIE(d) Balance in Tumour Angiogenesis. *Molecular Cancer Research* vol. 5 655–665 (2007).

48. Huang, H., Bhat, A., Woodnutt, G. & Lappe, R. Targeting the ANGPT–TIE2 pathway in malignancy. *Nat. Rev. Cancer***10**, 575–585 (2010).

Tables

Table 1. Properties of included microarray datasets in meta-analysis

Datasets / Reference	Patients No.		Stage I		Stage II		Stage III		Stage IV	Platform	Expression Array
	LUAD	Normal	A	B	A	B	A	B			
GSE75037	83	83	24	26	3	17	9	1	2	GPL6884	Illumina HumanWG-6 v3.0 expression beadchip
GSE63459	33	32	17	11	3	1	0	0	0	GPL6883	Illumina HumanRef-8 v3.0 expression beadchip
GSE32863	58	58	16	16	9	2	12	0	1	GPL6884	Illumina HumanWG-6 v3.0 expression beadchip
GSE10072	58	49	5	17	3	18	9	3	3	GPL96	[HG-U133A] Affymetrix Human Genome U133A Array
Total No.	232	222	62	70	18	38	30	4	6		

Table 2. The shared upregulated and downregulated DEGs among all stages of LUAD cases relative to normal lung tissues.

	Stages	DEGs
22 Upregulated DEGs for all stage	IA- control	ITPKA, NMU, MMP12, FUT3, ETV4, CP, COL1A1, PHLDA2, TK1, THBS2, STX1A, MELK, SPP1, CCNB2, SFN, SLC2A1, CDH3, CDC20, ASPM
	IIA- control	
	IB- control	
	IIB- control	
	IIIA- control	
	IIIB- control	
	IV- control	
140 Downregulated DEGs for all stage	IA- control	TEK,WIF1,DES,CRYAB,COX7A1,CACNA2D2,RASIP1,C7,TGFBR3,ZFP36,AOX1,FAM107A,VGLL3,C5AR1,TMEM204,LDB2,CD36,DPE CD93,GRK5,PID1,TSPAN7,TCEAL2,AQP4,JAM2,KANK3,CAV2,CPA3,PMP22,HSD17B6,CALCRL,ADRB2,SOSTDC1,MMRN1,CDH5,A FHL5,CLDN5,ABCA3,HBEGF,GNG11,TMEM100,PDK4,CDH13,RAMP2,KL,HIGD1B,THBD,WASF3,FAM189A2,SH3GL3,CA4,TCF21,C CDO1,GPC3,GHR,CLDN18,AGER,EMCN,RASL12,LRRN3,VEGFD,ID3,FMO2,PTGDS,FHL1,SPARCL1,STX11,GPX3,EPAS1,MS4A2,AD RAMP3,SFTPC,PCOLCE2,SLIT2,LAMP3,GPM6A,LIMS2,CAV1,MFAP4,ANGPT1,FABP4,VIPR1,CAVIN2,AOC3,CCL23,STXBP6,DNAS1 MT1M,EDNRB,CD52,CES1,LMO2,FCN3,LRRC36,CRTAC1,ALDH1A1,CLEC3B,ACADL,,P3H2,S1PR1,ACKR1,SRPX,LYVE1,GPM6B,FX LPL,FOXF1,CLIC5,TM6SF1,PECAM1,HBB,CPB2,LHFPL6,VSIG4,SOX17,FBLN5,ADAMTSL3,GDF10,RGCC,MARCO,PGC,PLA2G1B,H
	IIA- control	
	IB- control	
	IIB- control	
	IIIA- control	
	IIIB- control	
	IV- control	

Table 3: Unique upregulated and downregulated DEGs for each stage.

Stages	No	DEGs
Unique Upregulated DEGs for each stage	IA-control	3 CRLF1, TESC, AKR1B10
	IIA-control	8 CILP, ADAMDEC1, CD19, TMPRSS6, GFRA3, PLXNB3, CEMIP, BRINP1
	IB-control	1 FERMT1
	IIB-control	1 SIX1
	IIIA-control	12 SFXN1, COL1A2, CDH1, NPM3, PDIA4, KDELR2, TM4SF4, SLC35F2, CLDN10, SMPDL3B, ATAD2, RHBDL1
	IIIB-control	32 SPSB1, PCDH7, GSDMB, TGFA, KDELR3, MX2, PPP1R14D, XDH, PLEK2, FAP, DCBLD2, MST1R, INPP4B, SLC5A3, GOLM1, PLOD2, SRD5A3, DUOX2, FUT6, TRIM2, BATF, ABCC3, LAMC2, C1GALT1, TSTA3, PDZK1IP1, CPD, PAQR4, IL37, ITGA2, ENO2, HTR3A
	IV-control	84 NRIP3, BARD1, CCNE1, RAC3, DBNDD1, BOP1, SLC16A8, PCLAF, HIST2H2AA3, E2F2, TNNI3, ATIC, UCHL1, PTPRH, NEFH, ATP8B3, GALR2, AUNIP, ABCB6, NQO1, ST14, GALNT6, C17orf53, APOBEC3B, VGF, PRR11, FEN1, PKMYT1, CERS6, STRA6, POLQ, LRFN4, BIK, DNAJC22, EZH2, SLC25A10, GTSE1, IGFBP2, TTL12, GNG4, SKA1, HBQ1, AGT, SERPINB5, FANCI, QPRT, TNNI1, KRT19, TACC3, DNASE1, CENPA, MRPL12, RRM2, CLDN6, ASF1B, KPNA2, TMEM158, TIMP4, ETNK2, COCH, PPAT, ECEL1, FOXD2, EPN3, H1FO, MYO7A, CAD, RPL39L, CYP4F11, MC1R, KIF14, PRAME, YBX2, KIF1A, P3H4, HIST1H1C, MIF, CKS CENPM, COL7A1, LMNB2, HAMP, CTSV, SPC24
Unique Downregulated DEGs for each stage	IA-control	2 TMSB15A, MFAP5
	IIA-control	17 LDLR, AHNAK, FBP1, SCEL, COBL, EDN1, FOS, KRT4, MT1E, HLA-DQB1, PPARG, TBX2, MSLN, HLA-DQA1, FERMT2, PHACTR2, STO
	IB-control	0
	IIB-control	0
	IIIA-control	21 CCL2, KLRB1, DAPK2, GSTM3, GIMAP6, PENK, HCK, CCL21, SH3GL2, GSTA4, CCL3, GJC2, DENND2A, NKG7, STARD8, LY86, FCGR1
	IIIB-control	14 ID4, CCN2, RBP4, CDKN1C, BEX4, WFS1, NOTCH1, EML1, PRKD1, A2M, REV3L, SOBP, CADM1, SGK1
	IV-control	55 PIGR, ADGRF5, TPPP, FOXF2, MID1IP1, NUPR1, IRAK3, GLIPR1, SIK1, SLC02B1, ROS1, GLT8D2, ETS2, ITPR1, PAPSS2, PPP1R3C, IG KIT, GABARAPL1, PDLIM3, STON1, TRIM22, DRAM1, ENPP2, TCIM, B3GNT2, ALDH2, PTPN13, NDRG2, HOXA4, DPYSL2, HYAL1, RPRELP, SPON1, FGL2, HLA-DRA, C6, TMPRSS2, RGS2, IL18R1, SECISBP2L, PRR5L, MAN1C1, DOCK2, F13A1, FOLR1, MAOA, LIFR, CX3CR1, MAP3K8, PRRG1, S

Table 4. Top three co-expression clusters.

Cluster	Score (Density*#Nodes)	Nodes	Edges	Node IDs
1	6	6	15	CCNB2, CDC20, ASPM, MELK, TOP2A, TK1
2	6	6	15	GNG11, VIPR1, ADRB2, RAMP2, RAMP3, CALCRL
3	3.8	11	19	VWF, LYVE1, MMRN1, CFD, EMCN, CDH5, THBD, FIGF, PDK4, FABP4, LPL

Table 5: Metabolic pathways of top cluster genes.

Name	Source	FDR	Hit Count
G alpha (s) signalling events	REACTOME	1.56E-05	6
GPCR ligand binding	REACTOME	3.10E-02	5
Class B/2 (Secretin family receptors)	REACTOME	8.52E-04	4
Platelet activation, signaling and aggregation	REACTOME	3.69E-02	4
Calcitonin-like ligand receptors	REACTOME	4.20E-05	3
Platelet degranulation	REACTOME	3.69E-02	3
Response to elevated platelet cytosolic Ca2+	REACTOME	3.69E-02	3
Formation of Fibrin Clot (Clotting Cascade)	REACTOME	4.38E-02	2

Table 6. Metabolic pathways of shared genes

Name	Source	FDR	Hit Count
Hemostasis	REACTOME	3.62E-02	16
Extracellular matrix organization	REACTOME	2.08E-02	12
Platelet degranulation	REACTOME	2.92E-02	7
Response to elevated platelet cytosolic Ca2+	REACTOME	2.92E-02	7
Cell surface interactions at the vascular wall	REACTOME	3.14E-02	7
G alpha (s) signaling events	REACTOME	3.28E-02	7
Cell junction organization	REACTOME	2.78E-02	6
Phase 1 - Functionalization of compounds	REACTOME	3.62E-02	6
Cell-cell junction organization	REACTOME	2.92E-02	5
Defective B3GALT1 causes Peters-plus syndrome (PpS)	REACTOME	2.92E-02	4
O-glycosylation of TSR domain-containing proteins	REACTOME	3.14E-02	4
Erythrocytes take up oxygen and release carbon dioxide	REACTOME	2.08E-02	3
Calcitonin-like ligand receptors	REACTOME	2.08E-02	3
Erythrocytes take up carbon dioxide and release oxygen	REACTOME	2.78E-02	3
O2/CO2 exchange in erythrocytes	REACTOME	2.78E-02	3

Table 7. The top biological processes from gene ontology.

ID	Name	FDR	Hit Count
GO:0042127	regulation of cell population proliferation	1.27E-09	47
GO:0007155	cell adhesion	4.83E-09	40
GO:0022610	biological adhesion	4.83E-09	40
GO:0009719	response to endogenous stimulus	5.98E-06	38
GO:0051094	positive regulation of developmental process	1.12E-06	37
GO:1901700	response to oxygen-containing compound	1.43E-05	37
GO:0031226	intrinsic component of plasma membrane	1.24E-05	36
GO:0072359	circulatory system development	7.01E-08	35
GO:0051241	negative regulation of multicellular organismal process	1.47E-06	35
GO:0016477	cell migration	3.85E-05	35
GO:0046903	secretion	8.00E-05	35
GO:0051674	localization of cell	2.09E-04	35
GO:0048870	cell motility	2.09E-04	35
GO:0035295	tube development	1.05E-07	34
GO:0019220	regulation of phosphate metabolic process	6.31E-04	34
GO:0051174	regulation of phosphorus metabolic process	6.31E-04	34
GO:0042802	identical protein binding	4.01E-03	33
GO:0048646	anatomical structure formation involved in morphogenesis	1.75E-06	33
GO:0042325	regulation of phosphorylation	2.47E-04	33
GO:0006952	defense response	4.27E-04	33
GO:0001944	vasculature development	1.27E-09	32
GO:0035239	tube morphogenesis	1.51E-08	32
GO:0016050	vesicle organization	5.10E-04	32
GO:0005887	integral component of plasma membrane	1.05E-04	32

Figures

Figure 1
Heatmap of the shared 162 DEGs among all stages of LUAD cases relative to normal lung tissues. Row: gene; column: stage. Red: up-regulation; blue: down-regulation.

Co-expression network for common genes

Clusters: 8
Node: 38
Edge: 64
Score (Density Nodes): 3.85

Top 5 hub genes:
PECAM1, TEK, CDH5, VWF & ANGPT1

Figure 2

Supplementary Files

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

- [Supplementaryfile1.xlsx](#)
- [supplementaryfile1.rar](#)
- [Supplementaryfile1.zip](#)
- [supplementaryfile2.xlsx](#)
- [Supplementaryfile2.zip](#)
- [Supplementaryfile2.xlsx](#)