

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Identification of Key Genes Related to Lung Cancer with Brain Metastasis through Weighted Gene Correlation Network Analysis

Huatao Niu

Department of Neurosurgery[®]The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Honghua Cao

Department of Hematology, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Xin Liu

Department of Molecular diagnosis center, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Pin Zuo

Department of Neurosurgery®The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Yanbei Chen

Department of Neurosurgery[®]The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Zhaojin Cheng

Department of Neurosurgery[®]The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Ke Li (■ likelike@126.com)

Department of Cancer Biotherapy Center, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Yunchao Huang

Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center

Research Article

Keywords: Lung cancer, Brain metastases, WGCNA, Key disease genes

Posted Date: October 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-950975/v1

License: ©) This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background: Brain metastases (BM) are the most common malignant tumors affecting the human central nervous system, more than 50% of BM are lung cancer. However, there is currently no universal and effective screening tool for lung cancer with BM. The study aimed to identify key genes of lung cancer with BM, achieving early and accurate diagnosis of lung cancer with BM and long-term monitoring of the therapeutic response.

Methods: RNA-seq data and related clinical information were downloaded from the Gene Expression Omnibus (GEO) database. The weighted gene co-expression network (WGCNA) analysis was performed using WGCNA package and MEtuequois module was considered as the highest correlation. The Limma package of R was used to obtain differentially expressed genes (DEGs) between lung cancer with BM and healthy samples. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments of DEGs were analysed. LASSO and SVM were used to derive key genes for lung cancer with BM, and survival analysis was used to verify the correlation between the outcome and the key genes that we found. Key genes are annotated by GSEA, and CIBERSORT was used to calculate immune infiltration from samples.

Results: 5 co-expression modules were detected, and ME turquois was most significantly associated with BM in lung cancer. These genes were mainly enriched in biological processes related to cell adhesion and immunity, and the main signaling pathways are significantly related to disease and cell adhesion. 4 BM-related key genes (AQP1, GSTA1, ROS1, TGFB1I1) were identified by LASSO regression analysis and SVM-RFE screening, they were demonstrated as the potential markers for lung cancer with BM. Immune cell infiltration revealed that they were negatively correlated with T cells regulatory, T cells CD8 and Mast cells activated. Overall survival correlation analysis confirmed that the expression of AQP1, GSTA1 and TGFB1I1 was negatively correlated with the prognosis of lung cancer patients, ROS1 was positively correlated with the prognosis of patients.

Conclusions: AQP1, GSTA1, TGFB1I1 and ROS1 were identified as key genes of lung cancer with BM patients, which may regulate cancer progression by regulating immunity and have potential prognostic effects.

Background

Brain metastases (BM) are the most common malignant tumors that affect the human central nervous system and are also a terrible complication of systemic cancers throughout the body. It accounts for about 15%-30% of brain tumors, and its incidence is 10 times that of primary neuromas in the Brain [1]. It has been reported that over 50% of BMS are lung cancer BM, followed by melanoma, urogenital tumor and tumor of the digestive tract [2]. Moreover, patients with lung cancer BM have low quality of life, short survival period and poor prognosis [3]. Currently, the main treatment methods for BM include whole brain radiotherapy (WBRT), stereotactic radiosurgery (SRS), surgery and chemotherapy [4]. Sarah et al. found

that the median survival of patients who received WBRT after surgery was 40 weeks, which was improved compared with the radiotherapy group (15 weeks) [5]. Minniti et al. studied 66 patients with non-small cell lung cancer (NSCLC) BM treated with WBRT combined with SRS. The results indicated a median overall survival (OS) in the combined group was 10.3 months and in the WBRT group alone was 7.2 months (P=0.005) [6]. Currently, these therapeutic strategies are not ideal for the prognosis and survival of patients with lung cancer BM.

Early diagnosis and screening may be an important measure to improve the clinical treatment of lung cancer BM. However, there is currently no universal and effective screening tool for BM in lung cancer. Some studies have reported some biomarkers related to BMS, but the clinical effectiveness of these markers still needs to be further verified. For example, CEA, CA153, and cytokeratin 19 fragment had significantly better progression-free survival than patients with abnormally elevated lung cancer BM [7–9]. Matthew et al. found that p-S6 is overexpressed in metastatic tumors. In primary tumors, higher p-S6 expression is associated with shorter metastatic-free survival [10]. These results suggest that the level of tumor markers can be used as a factor in evaluating prognosis. In this study, the key genes related to lung cancer brain metastasis were screened out by WGCNA analysis based on GEO database, and the relationship between BMS related genes and tumor immunity was focused.

Materials And Methods

Data sources and processing

Three sets of data, GSE14108, GSE31547 and GSE126548, were downloaded from GEO, among which GSE14108 included 9 patients with lung cancer BM, GSE31547 included 30 patients with primary lung cancer, and GSE126548 included 6 patients with non-small cell lung cancer, of whom 3 had BM and 3 had no BM. TCGA database is used to download 504 lung squamous carcinoma samples.

Weighted gene correlation network analysis

In this study, hclustfunction was applied to cluster analysis of samples in the R WGCNA database, and samples with outliers removed were clustered according to gene expression levels to reveal the correlation between samples. The soft threshold power was determined according to the law of scale-free network, the gene cluster tree was constructed based on the correlation of gene expression levels, and the co-expression module was determined by dynamic pruning method. Modules with similar expression patterns are merged according to the similarity (0.4) of the module eigenvalues. The expression pattern of module genes in each sample was represented by the module eigenvalue, and the heat map of the expression pattern was drawn by the sample eigenvalue. Heat maps were used to identify modules that were significantly associated with the clinical features of BM.

Functional enrichment analysis

R's "SVA" plug-in was used for batch processing of GSE14108 data and GSE31547 data, and the Limma package was used to analyze differences in gene expression levels between primary lung cancer samples and BM groups (log2FC \geq 1, P \leq 0.05). Ggplot2 is used for heat map visualization. GO functional analysis and KEGG pathway analysis were performed on the module key genes using the R-package cluster Profiler software.

LASSO regression analysis and SVM –RFE screening

Key genes related to BM in lung cancer were further screened, R glmnet was used in the LASSO regression analysis of candidate feature genes. The radial basis function kernel SVM was independently adjusted for the 10-fold sampled data set, including the use of grid search for the SVM hyper parameters. Then, the optimal parameters are used to train the SVM over the entire data set, doing this for each fold in the external 10-fold CV, and averaging all of these 10 generalization error estimates to provide greater stability. Repeating this process, while changing the number of primary functions used as input, there is usually a "sweet spot" that minimizes the generalization error. According to this "best point", we can screen out our marker molecules to obtain key genes with diagnostic value.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) is used to evaluate the distribution trend of a pre-defined gene in the gene table ranked by the phenotype correlation, so as to judge its contribution to the phenotype. Lung cancer samples were divided into high and low expression groups according to the expression of each key gene, and the differentially expressed genes between high and low expression groups were enriched by GSEA pathway analysis. CIBERSORT was used to calculate 22 immune cell infiltrations in patients with BM / primary lung cancer, and Pearson correlation was used to analyze the association of key genes with immune cells.

Prognostic significance of key genes

The R Project for Statistical Computing's "Survival" package was used to analyze the correlation between the expression of key genes associated with BM and OS in lung cancer patients.

Results

WGCNA analysis results of modules and genes related to lung cancer brain metastasis

The Cluster tree displaying the relationship among the biological replicates of REPL samples was as shown in Fig. 1a, a total of 9 samples were clustered in BM sample, they are GSM354029, GSM354030,

GSM354025, GSM354027, GSM354032, GSM354028, GSM354033, GSM354026 and GSM354031. The soft-threshold power is a key value used to power the correlation of the genes and affects the mean connectivity and the scale independence of co-expression modules. When the soft-threshold power was β = 3, the scale independence was higher than 0.8 and the mean connectivity was higher (Fig. 1b). ME Diss Thres was set to 0.4 and combined with the dynamic clipping tree algorithm to split the module. After the cluster tree was generated, eight different gene co-expression modules were identified in the REPL, represented by different colors (Fig. 1c). According to the correlation between modules and sample traits (Fig. 1d), 5 modules were correlated with lung cancer BM (P \boxtimes 0.05), including ME Blue (Core = 0.59, P = 8E-5), ME Black (Core = -0.52, P = 7e-4), ME Brown (Core = 0.8, P = 7e-10), and ME Turquois (Core = 0.89, P = 4E-14) and ME Purple (Core = -0.4, P = 0.01) (Fig. 1e). Among them, ME turquois has the highest correlation with sample traits, showing a positive correlation, with a total of 5724 genes, of which 900 genes satisfy GS> 0.2 and MM> 0.8.

Screening, annotation and expression validation of brainmetastasis-related characteristic genes in lung cancer

After batch processing of sequencing data of GSE14108 and GSE31547, a total of 753 differentially expressed genes were obtained in BM and primary lung cancer samples. Among them, 238 genes were up-regulated in BM samples and 555 genes were down-regulated in BM tissues (Fig. 2a). Furthermore, the intersection of the differentially expressed genes of lung cancer brain metastasis samples and the module genes related to lung cancer brain metastasis resulted in 272 genes, which were used for subsequent analysis (Fig. 2b). As can be seen from the GO enrichment bubble diagram, cell adhesion and immune-related biological processes are the main pathways for the enrichment of key genes in the module, such as positive regulation of cell-cell adhesion, Humoral immune response and positive regulation of cell adhesion (Fig. 2c). The molecular function is mainly related to enzyme activity, including RNA polymerase II activating transcription factor binding, serine-type endopeptidase activity and metallopeptidase activity. As can be seen from the enrichment bar chart of KEGG pathway, the major participating signaling pathways are significantly related to diseases and cell adhesion, such as Pertussis, Rheumatoid arthritis, Malaria, Bladder cancer, Legionellosis, Influenza A, TNF signaling pathway, Focal adhesion, Complement and coagulation cascades and Age-RAGE signaling pathway in diabetic complications (Fig. 2d). The results of GSE126548 differentially expressed genes analysis showed that there were 773 differentially expressed genes between BM / primary lung cancer (Fig. 2e). Jvenn was used to intersect the differentially expressed genes of GSE126548 and the key genes of the module. A total of 14 intersection genes were obtained, indicating that these 14 genes may play an important role in lung cancer brain metastasis (Fig. 2f). The heat map results of further cluster analysis showed that the expressions of the 14 signature genes were significantly different in the transferred and primary samples (Fig. 2g).

LASSO regression analysis and SVM screening of characteristic genes in lung cancer brain metastases

In this study, LASSO regression analysis was used to screen for key genes associated with BM in lung cancer. When lambda. Min was minimal, the best diagnostic model was obtained, which contained 4 characteristic genes (AQP1, GSTA1, ROS1, and TGFB1I1) (Fig. 3a). The receiver operating characteristic (ROC) curve evaluates the accuracy and sensitivity of the diagnostic model, When the area under the ROC curve (AUC) value of the diagnostic model was 1, it indicated that the diagnostic model had high diagnostic value (Fig. 3b). SVM-RFE analysis found that when the number of genes changed from 1 to 14, the error rate of the optimal point for predicting BM lung cancer samples and primary lung cancer samples was 0.00865, and the accuracy rate was 0.991, which corresponding to 9 characteristic genes (ROS1, TGFB111, AEBP1, GSTA1, FMO3, AQP1, COMP, OLFM1, MMP1) (Fig. 3c). The intersection of characteristic genes screened by LASSO and SVM was considered as the key genes. The results showed that AQP1, GSTA1, ROS1 and TGFB111 may be the diagnostic biomarkers of BMS and primary lung cancer, and are the key genes to study the difference between primary lung cancer and BM (Fig. 3d).

Functional prediction of key genes in lung cancer brain metastases

According to the GSEA function prediction results of AQP1, GO biological processes such as immune response regulation and cell activator chromosome separation were significantly overexpressed in the high-expression group of AQP1. In addition, Cell adhesion, immunity and cytokine related pathways were significantly enriched in the group with high AQP1 expression (Fig. 4a). The GSEA function prediction results of GSTA1 showed that the GO biological processes such as cell division and chromosome separation were significantly low expressed in the high expression group of GSTA1. In addition, Cell cycle, disease-related pathways significantly enriched in high expression group (Fig. 4b). The GSEA function prediction results of ROS1 showed that the immune response was significantly high in the ROS1 high expression group. The prediction results of GSEA function of ROS1 showed that the immune response was significantly overexpressed in the group with high ROS1 expression, Cell division, chromosome separation and other GO biological processes were significantly low expressed in the high expression group of ROS1. In addition, cell cycle, disease infection, phagocytosis, immune-related pathways were significantly enriched in ROS1 high expression group (Fig. 4c). The GSEA function prediction results of TGFB1I1 showed that the biological processes such as immune regulation, lymphocyte activation and lymphocyte-mediated immunity were significantly enriched in the TGFB1I1 high expression group, while the GO biological processes such as cell division and chromosome separation were significantly decreased in the TGFB1I1 high expression group. In addition, pathways related to disease infection and phagocytes were significantly enriched in the TGFB1I1 high expression group (Fig. 4d).

Correlation between key genes of lung cancer brain metastases and immune invasion

The CIBERSORT algorithm and LM22 gene marker were used to analyze the proportion of 22 kinds of immune cells in the tumor samples. According to the calculation results of CIBERSORT, we drawn a heat map of the content of immune cells in each sample with P <0.05 (Fig. 5a). At the same time, Pearson correlation is used to analyze the relationship between immune cells. Fig. 5b shows the specific immune correlation coefficient. For example, the correlation coefficient between Plasma cells and B cells memory is -0.57 (Fig. 5c). Then the immune cell differences based on t-test were shown between BM / primary lung cancer tissues (Fig. 5d). Among them, immune cells B cells naive, Plasma cells, T cells CD8, T cells regulatory, Macrophages M1, Mast cells resting and Mast cells activated have significant differences. We further analyzed the correlation of 4 key genes and 7 immune cells with significant differences between the BM / primary lung cancer groups, and the results showed that the key genes were negatively correlated with T cells regulatory, T cells CD8 and Mast cells activated, and positively correlated with other cells (Fig. 5e).

Prognostic significance of key genes in brain metastases from lung cancer

Analysis of the correlation between the expression of key genes and the overall survival of lung cancer patients shows that AQP1, GSTA1 and TGFB1I1 were significantly correlated with the prognosis of patients with lung cancer, and the higher the gene expression, the worse the prognosis of patients. ROS1 is significantly correlated with the prognosis of lung cancer patients, and the higher the gene expression, the better the prognosis of patients (Fig. 6).

Discussion

WGCNA was applied because of its advantages in uncovering co-expression modules and their correlation with sample traits, and its higher reliability and biological significance compared with other bioinformatics approaches. Based on the WGCNA, LI et al. revealed novel regulatory modules associated with recurrent early pregnancy loss and Bao et al. identified the hub genes related to hypoxic adaptation in yak (*Bos grunniens*) [11, 12]. In this study, we identified 4 key genes associated with BMS in lung cancer based on WGCNA analysis, namely AQP1, GSTA1, TGFB1I1, and ROS1.

In this study, Aquaporin 1 (AQP1) was significantly associated with prognosis in patients with lung cancer. Kong et al. found that MicroRNA-133a-3p inhibits cell proliferation, migration and invasion in colorectal cancer by targeting AQP1 [13]. Angelico et al. found that Aquaporin-1 (AQP1) is highly expressed in asbestos-related malignant pleural mesothelioma (MPM) [14]. These studies all suggest that AQP1 is abnormally expressed in brain tumors and may play an important role in the invasion of tumors. In the functional prediction analysis, biological processes such as cell division and chromosome

separation were significantly low in the GSTA1 high expression group, while cell cycle and disease infection related pathways were significantly enriched in the GSTA1 high expression group. Glutathione S-transferase (GSTs), as the main phase II metabolic enzyme isozymes in cells, can catalyze the binding of glutathione to many electrophilic compounds, carcinogens, and some lipophilic anti-cancer drugs, protect cells from oxidative damage, and have anti-mutagenic and anti-mutagenic properties. It is believed that GSTs can slow cell proliferation by inhibiting signaling pathways. For example, GSTs inhibits JNK activity and has an important effect on JNK-mediated cell biological responses (proliferation, differentiation, apoptosis) [15, 16]. Aydin compared the percentage of positive staining between normal tissue and tumor tissue in patients with normal laryngeal cancer and patients with laryngeal cancer, and found that the expression of GSTA in normal cells was significantly higher than that in tumor cells (P⊠0.05) [17]. These studies suggest that GSTA1 has potential in tumor prognosis, which is consistent with our findings.

The GO enrichment results showed that cell adhesion and immune-related biological processes were the main biological processes for the enrichment of module key genes, and TGFB1I1 was significantly correlated with the prognosis of lung cancer patients. Previous studies have shown that transforming growth factor beta 1 induced transcript 1(TGFB1I1) can coordinate Rho GTPase activity to regulate adhesion dynamics and actin cytoskeleton remodeling during cell migration, thereby promoting tumor invasion and metastasis [18–21]. Nowadays, transcription factors such as Snail1, Slug, Zeb, Twist, βcatenin and Fox can affect the EMT process through the regulation of TGF-β, and TGFβ has been confirmed as the main inducer of tumor EMT. TGF-β can also promote the death of CD8 T cells by inhibiting the expression of B cell lymphoma2 (Bcl-2) [22]. At present, TGF-β and programmedcelldeath1 (pd-1) / programmedcelldeath1 ligand 1(Pd-I1) has the potential to improve T cell infiltration and alleviate the inhibitory function of PD-1 on T cells, and this therapeutic strategy is currently being validated in a variety of solid tumors [23]. In addition, we also found that reactive oxygen species (ROS) was significantly correlated with the prognosis of lung cancer patients. Under normal circumstances, ROS1 is activated after receiving specific information from the outside and transmits it to the inside of the cell to guide the cell to grow and proliferate in an orderly manner. If the ROS1 gene is mutated, it will cause cancer cells to drive up, which is the culprit of tumor occurrence. Glioblastoma [24], inflammatory myofibroblastoma [25, 26], cholangiocarcinoma [27], ovarian cancer [28], gastric cancer [29], colorectal cancer [30], angiosarcoma [31], arachnid melanoma [32] and non-small cell lung cancer [33] are all present with ROS1 fusion protein. Today, ROS1 is a proven therapeutic target for NSCLC.

Conclusion

In conclusion, we identified 4 key genes (AQP1, GSTA1, TGFB1I1 and ROS1) related with the progression of lung cancer BM by WGCNA analysis on DEGs database. We believe that these 4 genes may participate in lung cancer BM by regulating immune cell infiltration, which has potential prognostic effects. In other words, our findings may help predict, screen and early diagnosis of lung cancer BM patients. Additionally, the results will help identify therapeutic targets for lung cancer BM cancer and monitor the response to therapy.

Abbreviations

BM: brain metastases; WGCNA: weighted gene correlation network analysis; GEO: Gene Expression Omnibus; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; LASSO: least absolute shrinkage and selection operator; SVM-RFE: support vector machines-recursive feature elimination; WBRT: whole brain radiotherapy; SRS: stereotactic radiosurgery; NSCLC: non-small cell lung cancer; OS: overall survival; GSEA: Gene Set Enrichment Analysis; ROC: Receiver operating characteristic; AUC: Area under the ROC curve; AQP1:Aquaporin 1; GSTs: Glutathione S-transferase; TGFB1I1:transforming growth factor beta 1 induced transcript 1; ROS: reactive oxygen species.

Declarations Acknowledgements

Not applicable.

Authors' contributions

KL and YCH designed the research; PZ and ZJC collected and analyzed the data; HTN and HHC wrote the paper; XL and YBC revised the language of the paper. All authors read and approved the final manuscript.

Funding

This study was supported by Yunnan Fundamental Research Projects (202001AY070001-247);Reserve talents of Medical Science in Yunnan Province(H-2018055);Yunnan Education Department Project (2018JS229);Leader of medical Science in Yunnan Province (D-2017001)^{III}The 2015 Doctoral Scientific Research Fund Project of Yunnan Cancer Hospital (BSJJ201506).

Availability of data and materials

All the data were acquired from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

Consent for publication

This study did not involve the collection of patient samples.

Ethics approval and consent to participate

Clinical and animal experiments are not involved in this study. Therefore, ethical and participatory consent elements are not included.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Neurosurgery, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China.² Department of Hematology, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China.³ Department of Molecular diagnosis center, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China.⁴Department of Cancer Biotherapy Center, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center, No.519, Kunzhou Road, Kunming 650118, Yunnan, China.⁵ Department of Thoracic Surgery, The Third Affliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, No.519, Kunzhou Road, Kunming 650118, Yunnan, China.

References

- 1. Patchell, R.A., The management of brain metastases. Cancer Treat Rev, 2003. **29**(6): p. 533–40.
- 2. Pasqualini, R. and W. Arap, Profiling the molecular diversity of blood vessels. Cold Spring Harb Symp Quant Biol, 2002. **67**: p. 223–5.
- Nahed, B.V., et al., Congress of Neurological Surgeons Systematic Review and Evidence-Based Guidelines on the Role of Surgery in the Management of Adults With Metastatic Brain Tumors. Neurosurgery, 2019. 84(3): p. E152-e155.
- 4. Yousefi, M., et al., Lung cancer-associated brain metastasis: Molecular mechanisms and therapeutic options. Cell Oncol (Dordr), 2017. **40**(5): p. 419–441.
- 5. Goldberg, S.B., et al., Lung Cancer Brain Metastases. Cancer J, 2015. **21**(5): p. 398–403.
- 6. Minniti, G., et al., Stereotactic radiosurgery plus whole-brain radiotherapy for treatment of multiple metastases from non-small cell lung cancer. Anticancer Res, 2010. **30**(7): p. 3055–61.
- 7. Nieder, C., et al., Prognostic scores in brain metastases from breast cancer. BMC Cancer, 2009. 9: p. 105.
- 8. Louie, A.V., et al., Management and prognosis in synchronous solitary resected brain metastasis from non-small-cell lung cancer. Clin Lung Cancer, 2009. **10**(3): p. 174–9.

- 9. Sánchez de Cos, J., et al., Non-small cell lung cancer and silent brain metastasis. Survival and prognostic factors. Lung Cancer, 2009. **63**(1): p. 140–5.
- 10. McDonald, J.M., et al., Elevated phospho-S6 expression is associated with metastasis in adenocarcinoma of the lung. Clin Cancer Res, 2008. **14**(23): p. 7832–7.
- 11. Li, X., et al., Weighted gene correlation network analysis reveals novel regulatory modules associated with recurrent early pregnancy loss. Biosci Rep, 2020. **40**(6).
- Bao, Q., et al., Using weighted gene co-expression network analysis (WGCNA) to identify the hub genes related to hypoxic adaptation in yak (Bos grunniens). Genes Genomics, 2021. 43(10): p. 1231–1246.
- 13. Kong, B., et al., MicroRNA-133a-3p inhibits cell proliferation, migration and invasion in colorectal cancer by targeting AQP1. Oncol Lett, 2021. **22**(3): p. 649.
- 14. Angelico, G., et al., Immunohistochemical Expression of Aquaporin-1 in Fluoro-Edenite-Induced Malignant Mesothelioma: A Preliminary Report. Int J Mol Sci, 2018. **19**(3).
- 15. Wang, T., et al., Glutathione S-transferase P1-1 (GSTP1-1) inhibits c-Jun N-terminal kinase (JNK1) signaling through interaction with the C terminus. J Biol Chem, 2001. **276**(24): p. 20999–1003.
- 16. Liu, X., et al., Glutathione S-transferase Pi (Gstp) proteins regulate neuritogenesis in the developing cerebral cortex. Hum Mol Genet, 2021. **30**(1): p. 30–45.
- Sedat, A., et al., Immunohistochemical localization of glutathione s-transferase isoenzymes (gsta, Gstp, Gstm4, And Gstt1) and tumour marker p53 in matched tissue from normal larynx and laryngeal carcinoma: correlations with prognostic factors. J Otolaryngol Head Neck Surg, 2010. 39(5): p. 542–50.
- Shojaee, A., et al., Equine adipose mesenchymal stem cells (eq-ASCs) appear to have higher potential for migration and musculoskeletal differentiation. Res Vet Sci, 2019. 125: p. 235–243.
- Wang, H., et al., Nkx2-5 Regulates the Proliferation and Migration of H9c2 Cells. Med Sci Monit, 2020.
 26: p. e925388.
- 20. Ruan, X.J., et al., TGFβ1I1 suppressed cell migration and invasion in colorectal cancer by inhibiting the TGF-β pathway and EMT progress. Eur Rev Med Pharmacol Sci, 2020. **24**(13): p. 7294–7302.
- 21. Du, X., et al., HIC-5 in cancer-associated fibroblasts contributes to esophageal squamous cell carcinoma progression. Cell Death Dis, 2019. **10**(12): p. 873.
- 22. Sanjabi, S., M.M. Mosaheb, and R.A. Flavell, Opposing effects of TGF-beta and IL-15 cytokines control the number of short-lived effector CD8+ T cells. Immunity, 2009. **31**(1): p. 131–44.
- 23. Löffek, S., Transforming of the Tumor Microenvironment: Implications for TGF-β Inhibition in the Context of Immune-Checkpoint Therapy. J Oncol, 2018. **2018**: p. 9732939.
- 24. Stransky, N., et al., The landscape of kinase fusions in cancer. Nat Commun, 2014. 5: p. 4846.
- 25. Lovly, C.M., et al., Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. Cancer Discov, 2014. **4**(8): p. 889–95.

- 26. Yamamoto, H., et al., ALK, ROS1 and NTRK3 gene rearrangements in inflammatory myofibroblastic tumours. Histopathology, 2016. **69**(1): p. 72–83.
- 27. Marin, J.J.G., et al., Novel Pharmacological Options in the Treatment of Cholangiocarcinoma: Mechanisms of Resistance. Cancers (Basel), 2021. **13**(10).
- 28. Birch, A.H., et al., Chromosome 3 anomalies investigated by genome wide SNP analysis of benign, low malignant potential and low grade ovarian serous tumours. PLoS One, 2011. **6**(12): p. e28250.
- 29. Lee, J., et al., Identification of ROS1 rearrangement in gastric adenocarcinoma. Cancer, 2013. **119**(9): p. 1627–35.
- 30. Aisner, D.L., et al., ROS1 and ALK fusions in colorectal cancer, with evidence of intratumoral heterogeneity for molecular drivers. Mol Cancer Res, 2014. **12**(1): p. 111–8.
- 31. Giacomini, C.P., et al., Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. PLoS Genet, 2013. **9**(4): p. e1003464.
- 32. Wiesner, T., et al., Kinase fusions are frequent in Spitz tumours and spitzoid melanomas. Nat Commun, 2014. **5**: p. 3116.
- 33. Shen, P.W., et al., Disruption of Cytosolic Folate Integrity Aggravates Resistance to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors and Modulates Metastatic Properties in Non-Small-Cell Lung Cancer Cells. Int J Mol Sci, 2021. 22(16).



WGCNA screening of lung cancer BM-related modules and genes. a Sample dendrogram and trait heatmap. Branches represent samples, and the ordinate represents the height of hierarchical clustering. b The graph shows the scale-free fit index for various soft threshold powers to identify the optimal soft threshold power (β). In the graph on the left, the horizontal axis represents the soft threshold power or β values and the vertical axis represents the scale-free network index (R2). The scale-free characteristics of

the gene network are stronger when the R2 value is higher. In the right graph, the horizontal axis represents the soft threshold power or β values, the vertical axis represents the means of all the gene adjacency functions in the corresponding gene module. c Cluster dendrogram of all filtered genes enriched based on the dissimilarity measure and the cluster module colors. d module and trait heat map. e: The scatter plot shows the correlation between module genes and traits.



Screening, annotation and expression validation of brain-metastasis-related characteristic genes in lung cancer. a The volcano plot shows the expression profiles of 753 genes in BM samples compared to normal liver samples from the TGCA database. The low expressing genes (n=555) are shown in green and the high expressing genes (n=238) are shown in red. The threshold criteria are P=0.05 and log2FC=1. b Screening of key genes of module, green represents genes that are differentially expressed between BM / primary (n=5724), and blue represents WGCNA module genes related to BM (n=733).c The bubble map was used to show the enrichment results of TOP 6 in Biological process (BP), Cellular component (CC) and Molecular function (MF) directories, the size of the circle represents the number of differentially expressed genes, and the color from blue to red indicates that the credibility of the result is from low to high. d KEGG pathway analysis. The length of the bar indicates the number of differentially expressed genes enriched in the pathway, and the color from blue to red indicates that the credibility of the result is from low to high. e: The volcano plot shows the expression profiles of 773 genes in GSE126548 BM samples compared to normal liver samples from the TGCA database. The threshold criteria are P=0.05 and log2FC=1. f: Screening of key genes of module, the green module represents the key genes, and the blue represents the differentially expressed genes of GSE126548. g: Heat map of candidate characteristic gene expression between BM / primary group, the darker the color (red is up-regulation, blue is downregulation). Each column represents the expression level of all genes in each sample.



LASSO and SVM analysis of candidate feature genes. a Characteristic genes were screened by LASSO regression analysis. The abscissa deviance represents the proportion of residuals explained by the model, showing the relationship between the number of characteristic genes and the proportion of explained residuals. The ordinate is the coefficient of the gene (left). The abscissa is log (Lambda), and the ordinate represents the cross-validation error (right). b The diagnostic model was evaluated by the ROC curve, AUC refers to the surface area under the ROC curve. c SVM characteristics and error rate and accuracy rate. d The intersection of LASSO characteristic gene and SVM characteristic gene.



GSEA function prediction of key genes. a GSEA enrichment results of AQP1, the left side represents GOBP pathway and the right side represents the KEGG pathway. b GSEA function enrichment results of GSTA1, the left side represents the GOBP pathway and the right side represents the KEGG pathway. c The GSEA function enrichment results of ROS1, the left side represents the GOBP pathway and the right side represents the KEGG pathway. d GSEA enrichment results of TGFB1I1, the left representing GOBP pathway and the right side represents the KEGG pathway.



Immune infiltration status and correlation analysis in patients with BM and primary lung cancer. a The proportion of immune cells in tumor tissue. b The correlation heat map of the proportion of immune cells. The color indicates the correlation size, red indicates positive correlation and blue indicates negative correlation. The darker the color, the stronger the positive / negative correlation. c Significance test of correlation of the proportion of immune cells, P <0.05 indicates significant correlation. d Box plot of

different immune cells between BM / primary lung cancer, blue indicates the brain metastasis group, and yellow indicates the primary group. "*" means P <0.05, "**" means P <0.01, "***" means P <0.001, and "****" means P <0.0001. e Correlation diagram of key genes and differential immune cells, blue indicates positive correlation, yellow indicates negative correlation.





Figure 6

The survival curves of AQP1, GSTA1, TGFB1I1 and ROS1. Red represents the high expression group and blue represents the low expression group.