AGRP and ESPL1 as Biomarkers of Brain-Metastasis in Lung Adenocarcinoma

Jianzhi Deng¹, Xiaohui Cheng¹ and Yuehan Zhou²,*

¹Guilin University of Technology, Guilin 541004, China
²Guilin Medical University, Guilin 541004, Guangxi, China
Email: dengjianzhi@163.com
Email: yuehanzhou2012@163.com
*Corresponding Author

Abstract. In the mortal diseases, lung cancer, approximately one fifth of lung adenocarcinoma (LUAD) patients are associated with brain metastasis (BM). And even some patients die of BM. In the present research, the 22753 genes data of 273 primary LUAD or BM samples were downloaded from gene expression omnibus (GEO) datasets. The 145 common differentially expressed genes (DEGs) both from GPL96 and GPL570 platform profiles were screened out by R package. Gene ontology (GO), pathway and protein protein interaction (PPI) network analysis of the DEGs was enriched by online tools. The 14 up- and 131 down-regulated genes were enriched in 18 GO terms and 147 signal pathway. Protein expression of AGRP and ESPL1 is lower in lung cancer than many other cancers. And the km-plots of lung cancer survival curves are also shown that AGRP and ESPL1 express higher in high risk groups. From the research in this paper, we can believe that AGRP and ESPL1 might be the biomarkers for the diagnostic clue of BM from LUAD.

1. Introduction

Lung cancer is one of the highest mortal diseases worldwide. About 40-50% of lung cancer cases are diagnosed as lung adenocarcinoma (LUAD). During the course of diseases, approximately 16% of LUAD patients are associated with brain metastasis (BM) [1-4]. Even the high mortality of LUAD patients is due to BM. CT and MRI are the main methods of BM diagnosis, but the biomarkers that predict BM from LUAD are rarely reported. If we could find out the biomarkers of BM, it could help us to confirm the diseases and the best strategy for targeted therapy.

In this study, we aimed to analysis the BM related DEGs and its biological process (BP), cellular component (CC), molecular function (MF). First, we got nine gene expression profiles from GEO, and screened 145 significant DEGs between BM from LUAD and primary LUAD by R software. Secondly, GO enrichment and pathway of the significant DEGs were analysed by DAVID and KOBAS separately [5, 6]. At the end, AGRP and ESPL1, were researched as the biomarkers by CCLE [7], km-plot [8] and STRING [9].

2. Materials and Methods

2.1. Microarray Data Download and Processing

The microarray profiles, GSM461786, GSM461788 and GSM461790[10], GSE31548[11], GSE10072[12], GSE14108[13], GSE27716[14], GSE31546[11], GSE40791[15], were downloaded from the Gene Expression Omnibus (GEO) database. They were generated on GPL96 or GPL570. The BM homo sapiens organism samples were obtained from GSE14108, GSM461786, GSM461788 and...
GSM461790, while the primary LUAD homo sapiens organism samples were obtained from GSE31548, GSE10072, GSE27716, GSE31546 and GSE40791. The profile information is shown in table 1 and table 2. Each profile is contained 22753 genes expression.

Table 1. Information of GPL96 platform microarray profiles

| GEO      | LUAD | BM               | Sample NUM |
|----------|------|------------------|------------|
| GSE31548 | Yes  | Not mention      | 30         |
| GSE10072 | Yes  | Not mention      | 58         |
| GSE14108 | Yes  | yes              | 9          |

Table 2. Information of GPL570 platform microarray profiles

| GEO      | LUAD | BM               | Sample NUM |
|----------|------|------------------|------------|
| GSE27716 | Yes  | Not mention      | 40         |
| GSE31546 | Yes  | Not mention      | 17         |
| GSE40791 | Yes  | Not mention      | 94         |
| GSE14108 | Yes  | Yes              | 19         |
| GSE18549 | Yes  | Yes              | 3          |
| GSM461786| Yes  | Yes              | 1          |
| GSM461788| Yes  | Yes              | 1          |
| GSM461790| Yes  | Yes              | 1          |

Genes in the downloaded microarray profiles were represented as Affymetrix platform probe ID. And the probe ID was converted to the official gene symbol by using annotation file (ftp://ftp.ncbi.nih.gov/geo/platforms) and gene2accession file (ftp://ftp.ncbi.nih.gov/gene/DATA) from NCBI website.

2.2. Differentially Expressed Genes Screening and Cluster Analyzing

The microarray data from different profiles were separated into 2 groups according to the different platform GPL96 and GPL570. In each group, the microarray data were divided into metastasis sub-group and primary sub-group. DEGs from the metastasis sub-group and primary sub-group were screened by the limma and impute package of R software[16]. According to platforms GPL96 and GPL570, it was obtained 2 groups of DEGs respectively. The common DEGs of GPL96 group and GPL570 group were selected by using venn method for further research. Hierarchical clustering heatmaps of the common DEGs were screened by using heatmap R package. The DEGs screening and clustering were analyzed with the thresholds: corrected P-value<0.05 and |LogFoldChange|> 2[17].

2.3. DEGs GO Functional and Pathway Analyzing

GO enrichment analysis for DEGs is accomplished by DAVID, the Database for Annotation, Visualization and Integrated Discovery (https://david.ncifcrf.gov/). In GO enrichment analysis, we focus on the three function groups, including biological process, molecular function and cellular component with the thresholds: corrected P-value<0.05 and gene number>10. Relationships of the common DEGs and GO terms were analyzed by GOplot R package and were also shown in chordplot(count>10, ease<0.05). In pathway enrichment analysis, the screened DEGs were analyzed by onlne tool, KOBAS (kobas.cbi.pku.edu.cn).

2.4. CCLE, Kmplot, PPI Analyzing and Biomarker Screening

Cancer Cell Line Encyclopedia (CCLE) was an online tool (portals.broadinstitute.org), which was used to visual display the distribution of a gene’s expression in different tumor organism. Average distribution of the selected gene's expression was sorted and colored in the boxplot. The highest one was display on the left and was colored red. The biomarkers were screening by CCLE from the screening common up-regulated DEGs. The gene lower distribution in the boxplot was regarded as the biomarker. Clinical manifestation of the remarkable biomarkers, were confirmed by KM-plot of 1926
clinical specimens (www.kmplot.com). PPI was used to analyse the interaction between the DEGs by STRING (string-db.org).

3. Results

3.1. DEGs Screening from the GEO Datasets/ DEGs Clustering Analysis

After extracting and screening the DEGs of primary LUAD samples and BM samples from the selected microarray profiles, 557 DEGs were screened from the GPL96 group. Additionally, 720 DEGs were screened from the GPL570 group. There were 145 common genes both expressed differentially in GPL96 and GPL570 group. Among them, 14 DEGs were up-regulated, and the left 131 DEGs were down-regulated. The most significant DEGs (about 20% of the DEGs), including 14 up-regulated genes and 14 down-regulated genes.

Table 3. DEGs both significantly different expressed in GPL96 and GPL570

| DEGs                                | Gene symbol                   |
|-------------------------------------|--------------------------------|
| up-regulated                        | EEF1A2, PLP1, CTNNB1, UBE2S, EIF2S3, EIF5A, SNCG, UCHL1, AGRP, ITPKA, CEP55, TGFAP, ESPL1 |
| down-regulated                      | DNAJC12, EDNRB, MST1L, TRIM5, WNT5A, NFE2L3, CCDC68, BDH2, CCR7, SLC34A2, CASP1, SLC24A3, CX3CR1, MITF, GPR65, APOBEC3A, SLC15A2, LGR4, LTB, DMBT1, SDRP, EVI2B, GSAP, DRAM1, SLAMF7, TT2C8, CLEC2B, SRPX, SAMHD1, GJA1, IGSF6, COL4A4, FABP4, ENPP2, DPT, COLEC12, SOCS2, ITK, SAMSN1, ADAMDEC1, HSD11B1, LAMP3, CNN1, TMEM100, TNFRSF17, PLN, HLADMB, TLRL3, TMEM47, CLEC3B, NRARA2, FMO3, IL7R, CEACAM6, GNPNMB, IGHM, HLADPA1, TCF21, ALDH1A1, GIMAP4, FGL2, GPBP1, PROS1, ANOS1, BLNK, LY75, ROS1, FAT4, CXCL9, Mnda, PTPRC, FBN1, PDLI1, M3, GIMAP6, C4BP, SFRP4, BCL2A1, GPR183, GPR171, FOXF1, CXCL2, MOX1D1, P2RY14, CTSE, PLA2G7, GATA6, UBD, CXCL13, RRAD, PTPN13, IGLJ3, CHIT1, PLA2G1B, CYTIP, ASPN, GAS1, COMP, IL33, NR3C2, GREML1, IL6, GSTA2, RASGRP1, CPA3, MMP12, AQP1, MMP7, EGFL6, FLRT3, POU2AF1, LTF, ADH1B, CD69, EGR2, CD52, CCL18, CYP4B1, CCL19, ITGB1L1, VCAM1, CLIC5, LUM, HSD17B6, FMO2, MMP2, RARRES2, SFTPC, SCGB1A1, AMY1A, SFTPD, AOC3 |

Figure 1. Chord plot of DEGs

3.2. DEGs GO and Pathway Enrichment Analysis

The 145 common DEGs were enriched in 18 Go terms. In BP group, the DEGs were mainly enriched in immune response, defense response to virus, inflammatory response, signal transduction, cell
adhesion, proteolysis, innate immune response, negative regulation of apoptotic process and G-protein coupled receptor signaling pathway. In CC group, the DEGs were mainly enriched in extracellular space, extracellular region, proteinaceous extracellular matrix, cell surface, integral component of plasma membrane, extracellular exosome, endoplasmic reticulum membrane and plasma membrane. And the DEGs were enriched in calcium ion binding in MF group. The 10% most significantly up-regulated or down-regulated genes enriched in 5 GO terms shown in Fig.1. In pathway enrichment analysis(P-value<0.05), the 145 common DEGs were enriched in 147 pathways, especially in Cytokine-cytokine receptor interaction(hsa04060), Metabolic pathways(hsa01100), Chemokine signaling pathway(hsa04062), Pathways in cancer(hsa05200) and HTLV-I infection(hsa05166). The highest pathway enrichment analysis diagrams are shown in Fig.2.

3.3. Biomarker Screening
The biomarkers was screening by CCLE from the 14 common up-regulated DEGs firstly. And it can be known from the boxplot in Fig.3, the distribution of AGRP and ESPL1 were lower in LUAD(lung_NSC in Fig.3). As shown in Fig.4, the prognostic values of AGRP and ESPL1 were determined by KM plotter. Survival curves were drafted for all 1,926 LUAD patients (Fig.4(a)), and the KM-plot curves showed that high mRNA expression of AGRP was correlated with worse OS (HR=1.14,1.01-1.29, logrank P=0.041), and high mRNA expression of ESPL1(Fig.4(b)) also tend to worse OS of LUAD (HR=1.57,1.38-1.79, logrank P=2.2e-12).
4. Discussion

BM is a common complication in LUAD patients. And in many case, a targeted medicine for primary LUAD like Gefitinib, is not available in BM treatment. So we need to find out the unique DEGs for BM diagnose and therapy. In this paper, the microarray profiles, GSE31548, GSE10072, GSE14108, GSE27716, GSE31546, GSE40791, GSM461786, GSM461788 and GSM461790, were downloaded from the GEO database that is generated on GPL96 or GPL570. There are 14 up-regulated and 131 down-regulated DEGs between BM and primary LUAD were screened from the datasets above. The DEGs is mainly enriched in 18 GO terms, such as extracellular space, extracellular exosome, plasma membrane, immune response, integral component of plasma membrane, signal transduction and proteolysis. The DEGs is mainly enriched in 147 pathways, such as Cytokine-cytokine receptor interaction, Metabolic pathways, Chemokine signaling pathway, Pathways in cancer and HTLV-I infection. Among then, AGRP(Agouti-related protein) and ESPL1(extra spindle pole bodies like 1) are two typical DEGs in BM. AGRP was enriched in extracellular space(GO:0005615), extracellular exosome(GO:0070062) and Adipocytokine signaling pathway(hsa04920). ESPL1 was enriched in proteolysis (GO: 0006508), pathway Oocyte meiosis(hsa04114) and Cell cycle(hsa04110).And the Protein Interactions of the 14 up-regulated DEGs were shown in Fig5.

AGRP and ESPL1 are high expressed in BM rather than in primary LUAD (Fig.3 and Fig.4). And the km-plots of AGRP or ESPL1 analysis from 1926 patients were also shown poor prognosis. So, we can believe that the death of patients maybe cause by BM. And AGRP and ESPL1 can be regard as the important biomarkers of BM from LUAD for the diagnostic criterion.
5. Acknowledgments
This work was supported by the Guangxi Natural Science Foundation of China (2014GXNSFBA118151), National Natural Science Foundation of China (No. 81660031, 81360090), and Guangxi key Laboratory Foundation of Embedded Technology and Intelligent System.

6. References
[1] R. Komaki, J. D. Cox, and R. Stark, Frequency of brain metastasis in adenocarcinoma and large cell carcinoma of the lung: correlation with survival, Int J Radiat Oncol Biol Phys, vol. 9, no. 10, pp. 1467-70, Oct 1983.
[2] Y. Yuan et al., Activity of pemetrexed and high-dose gefitinib in an EGFR-mutated lung adenocarcinoma with brain and leptomeningeal metastasis after response to gefitinib, World J Surg Oncol, vol. 10, p. 235, Nov 7 2012.
[3] C. Ye, J. Wang, S. Zheng, and Y. Chai, Effective treatment with icotinib in lung adenocarcinoma with EGFR and ALK co-alterations and brain metastasis, Onco Targets Ther, vol. 9, pp. 6605-6608, 2016.
[4] K. Jayashree, C. Anubuti, Sunila, and M. Gundappa, Primary fallopian tube adenocarcinoma with brain and lung metastasis, Indian J Pathol Microbiol, vol. 52, no. 4, pp. 596-8, Oct-Dec 2009.
[5] C. Xie et al., KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases, Nucleic Acids Res, vol. 39, no. Web Server issue, pp. W316-22, Jul 2011.
[6] W. Huang da, B. T. Sherman, and R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic Acids Res, vol. 37, no. 1, pp. 1-13, Jan 2009.
[7] J. Barretina et al., The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature, vol. 483, no. 7391, pp. 603-7, Mar 28 2012.
[8] Jianzhi Deng, Xiaohui Cheng, Yuehan Zhou, PRLR is a target of oncogenic miR-204 in Low Grade Glioma and involved in a ceRNA network by Computational Biology, presented at the AMMSO2019, guilin, 2019.4.22, 2019.
[9] D. Szklarczyk et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res, vol. 47, no. D1, pp. D607-D613, Jan 8 2019.
[10] Jianzhi Deng, Xiaohui Cheng, Yuehan Zhou, Gene clustering, enrichment and survival analysis of differentially expressed genes in Low Grade Glioma between different genders by big data analysis, presented at the ICDMML2019, Hong Kong, 2019.4.28, 2019.
[11] Primary Lung Cancer Specimens, ed, 2011.
[12] M. T. Landi et al., Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival, PLoS One, vol. 3, no. 2, p. e1651, Feb 20 2008.
[13] F. Luke et al., Isolated metastasis of an EGFR-L858R-mutated NSCLC of the meninges: the potential impact of CXCL12/CXCR4 axis in EGFRmut NSCLC in diagnosis, follow-up and treatment, Oncotarget, vol. 9, no. 27, pp. 18844-18857, Apr 10 2018.
[14] A. C. Borczuk et al., Progression of human bronchioalveolar carcinoma to invasive adenocarcinoma is modeled in a transgenic mouse model of K-ras-induced lung cancer by loss of the TGF-beta type II receptor, Cancer Res, vol. 71, no. 21, pp. 6665-75, Nov 1 2011.
[15] Y. Zhang et al., USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis, J Clin Invest, vol. 122, no. 12, pp. 4362-74, Dec 2012.
[16] L. Heng et al., Molecular characterization of metastatic osteosarcoma: Differentially expressed genes, transcription factors and microRNAs, Mol Med Rep, vol. 15, no. 5, pp. 2829-2836, May 2017.
[17] P. Wang et al., Identification of biomarkers for the detection of early stage lung adenocarcinoma by microarray profiling of long noncoding RNAs, Lung Cancer, vol. 88, no. 2, pp. 147-53, May 2015.