Original Article
Bioinformatic analysis of the role of MNAT1 in the progression of lung cancer and its correlation with the prognosis of patients

Lifei Wang[1], Furong Tan[1], Yali Qiu[1], Gang Chen[2]

[1]Department of Respiration, The Third Hospital of Changzhou, Changzhou, China; [2]Department of Respiration, The Third Hospital of Hebei Medical University, Shijiazhaung, China

Abstract: Objective: The objective of the present study was to analyse the effects of MNAT1 on the progression of lung cancer and its role in predicting the prognosis of patients with lung cancer. Methods: Gene expression array data of lung cancer patients and non-lung cancer patients was downloaded from a public database (GEO) for bioinformatics analysis. DAVID and STRING databases were used to screen differentially expressed genes, which were then subjected to GO enrichment analysis, signal pathway analysis and protein interaction analysis. In addition, the clinical data of the lung cancer patients was obtained from the Oncomine database and used to analyse the effect of MNAT1 expression levels on overall survival. Results: A total of 178 differentially expressed genes were obtained from the matrix database, among which 14 genes were expressed at a high level, including MNAT1, and 164 were expressed at a low level. GO enrichment analysis showed that the differentially expressed genes were mainly located in the nucleus and cytoplasm and were involved in cell cycle regulation and DNA repair. KEGG signalling pathway analysis showed that the signalling pathways related to differential gene aggregation were basal excision repair and nucleotide excision repair. The protein interaction network showed that the main related proteins included MNAT1, SUPT5H, TDRD12, and TRIM21. Oncomine data analysis showed that MNAT1 was highly expressed in lung cancer patients, and the survival rate of patients with a high MNAT1 expression was lower than those with a low MNAT1 expression. Conclusion: High expression of MNAT1 predicts a poor prognosis in lung cancer patients, and MNAT1 might function as an oncogene in the development of lung cancer.

Keywords: Lung cancer, MNAT1, bioinformatics, GEO, survival rate

Introduction

For the Global Cancer(GLOBOCAN) standard [1], developed by the International Agency for Research on Cancer in 2018, researchers conducted a survey of data on 36 cancers in 185 countries. It was estimated that 18.1 million new cancer cases and 9.6 million deaths would occur in 2018. These data indicate that cancer has a serious impact on human life and health worldwide. The three most common cancers are lung cancer (11.6%), breast cancer (11.6%), and prostate cancer (7.1%). The cancers with the highest mortality rates are lung cancer (18.4%), colorectal cancer (9.2%), gastric cancer (8%), and liver cancer (8.2%). These results indicate that the incidence and mortality of lung cancer are the most prevalent, so the active diagnosis of early lung cancer plays a key role in the prognosis and survival of patients. The main causes of lung cancer are [2] smoking, genetics, environmental pollution, occupational exposure, diet, genetic mutations [3], etc. Research on lung cancer for gene mutations has gradually become a research hotspot, and some genetic monitoring methods and treatment methods have been applied in the clinic [4]. This study focused on the high-throughput sequencing of lung cancer patients and normal lung tissue to analyse the differential gene expression in the two groups, i.e., differential gene biological functions, signalling pathways, and selection of typical genes for lung cancer patients. Survival analysis was performed to observe the role of the MNAT1 gene in lung cancer and further assist in clinical treatment.

Materials and methods

Material

Bioinformatics data GSE54495 ID: 25705890; data was retrieved from the NCBI (Gene Expression Omnibus, GEO) database. The data research type was expression profiling by an array with chip platform GPL570, and the species was Homo sapiens. The analysis included data from 17 lung cancer patients and 13 non-lung cancer patients.

Data processing and differential gene analysis

Data processing were performed on the original data set using R software. The background correction, standardization and expression values of the original data were calculated using the RMA algorithm. Differentially expressed genes that were screened needed to meet $P < 0.05$ and Log $2 \geq 1$.

Bioinformatics analysis of differentially expressed genes

Database for annotation, visualization and integrated discovery (DAVID) is an online (https://david.ncifcrf.gov/) bioinformatics tool that synthesizes the information of a large number of genes and proteins. It also performs biometrics feature annotations. Differential genes were uploaded, and Gene Ontology enrichment (GO) [4] and Kyoto encyclopaedia of genes and genomes (KEGG) [5] pathway enrichment analysis were performed. $P < 0.05$ and FDR < 0.05 were set as having a critical value of significant gene enrichment.

Interaction analysis of differently expressed genes

STRING 10.0 (Search Tool for the Retrieval of Interacting Genes/Proteins) [6] database was formed using known or predicted protein interaction data, which includes the analysis of interactions between direct and indirect proteins. It was used to find protein interactions and finally, a network analysis of the protein-protein interactions (PPI) was constructed using Cytoscape software [7]. The data setting condition was a score >0.4.

MNAT1 expression analysis

To explore the expression patterns of MNAT1 in lung cancer patients and non-lung cancer patients and the effect of MNAT1 expression on patient survival, the Oncomine database was used.

Statistics

Quality control involves processing the file, from the simplest direct observation to applying commonly used methods and then more advanced data fitting methods [8]. The differentially expressed genes were analysed by t-test and the SAM method [9]. Hypergeometric testing was used for the GO enrichment and KEGG pathway analyses [10]. Differences were considered statistically significant at $P < 0.05$.
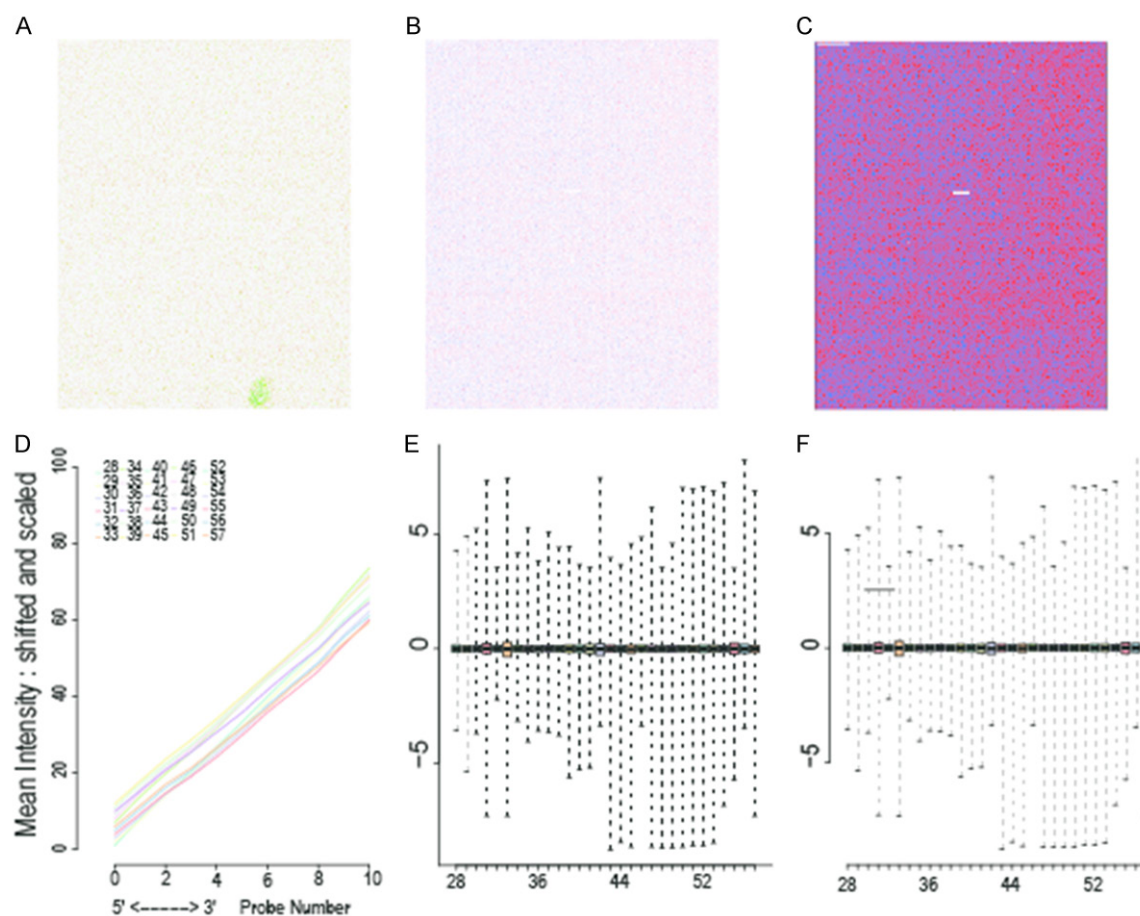
Results

Exploratory data analysis

Requirements for quality control of the downloaded data were as follows: 1. The median value of the gene changes at least 2 times; 2. The difference in gene expression requires $P < 0.02$; and 3. The amount of missing values of the data must not be < 50%. Quality control of the data showed that the RLE was in the same horizontal line and the RNA degradation map, the weight map, the weight symbol map, and the residual map showed uniform images, indicating that the stability and quality of the specimen were in a good range and that the data were analysable (**Figure 1**).

Screening of differently expressed genes

High-throughput sequencing analysis of 13 non-lung cancer patients and 17 lung cancer patients showed that there were 178 significant differential genes (the screening of differential genes needed to meet $P < 0.05$ and Log $2 \geq 1$). Fourteen of the differentially expressed genes were expressed at high levels (LOC102723448, ITGA9-AS1, NUPR1, IFI35, IFI35, KIRREL3, MNAT1, FOLR3, FOLR3, CA3, MMP12, SCG2, SCG2, ALDH1A2), and 164 of the differentially expressed genes were expressed at low levels (these mainly include SLC25A1, PPP1R15B, LOC100507291, WBSCR27, LIPT1, CLDN10, NEURL4, PSMB3, AGL, FBXO5, KIAA0895, EIF2D, and LIMS2). R language was used to visualize the differential genes as heat maps (**Figure 2**), green indicates a low expression and red indicates a high expression). **Figure 3** shows a volcano plot.

Figure 1. Data Quality Assessment (A-C: Weight residual plot of the GSE54495 dataset; D: RNA degradation profile of the GSE54495 dataset; E: NUSE box plot of the GSE54495 dataset; F: RLE box plot of the GSE54495 dataset).

GO analysis results

The DAVID website was used to perform GO enrichment analysis on differential genes, and the results showed that the differential genes of the GSE54495 data set are mainly located in the nucleus, cytoplasm, and telomere region, where their main roles involve regulating neural tube development and RAC protein signal transduction. Other roles include negative regulation, DNA-template transcription, prolongation, dephosphorylation of methylguanosine mRNA, transcriptional extension of RNA polymerase II promoter, gene excision repair, DNA repair, negative regulation of hepatocyte proliferation, translation initiation dependent on IRES, ribosome disintegration, DNA glycosylase activity, cell cycle, DNA template transcription, positive regulation of elongation, positive regulation of triglyceride biosynthesis, impaired DNA binding, DNA (apurinic or apyrimidinic)-lytic enzyme activity, protein phosphatase type I complex, transcription-coupled nucleic acid-excision repair, etc. (**Table 1** and **Figure 4**).

KEGG signal pathway analysis results

Signal pathways of the differential genes were obtained by KEGG analysis. The main aggregation pathways were basal excision repair, nucleotide excision repair, cell adhesion molecules (CAMs), the complement cascade and the coagulation cascade. The signalling pathway of the MNAT1 gene are shown in **Table 2** and **Figure 5**.

Module analysis of protein interactions

The STRING website screens data to determine protein interactions. The proteins of the top 10 core genes were screened and mainly MNAT1, SUPT5H, TDRD12, TRIM21, SLC25A1, TTC16, LRRC4, RALYL, OXNAD1, DCLRE1A were found (**Figure 6**).
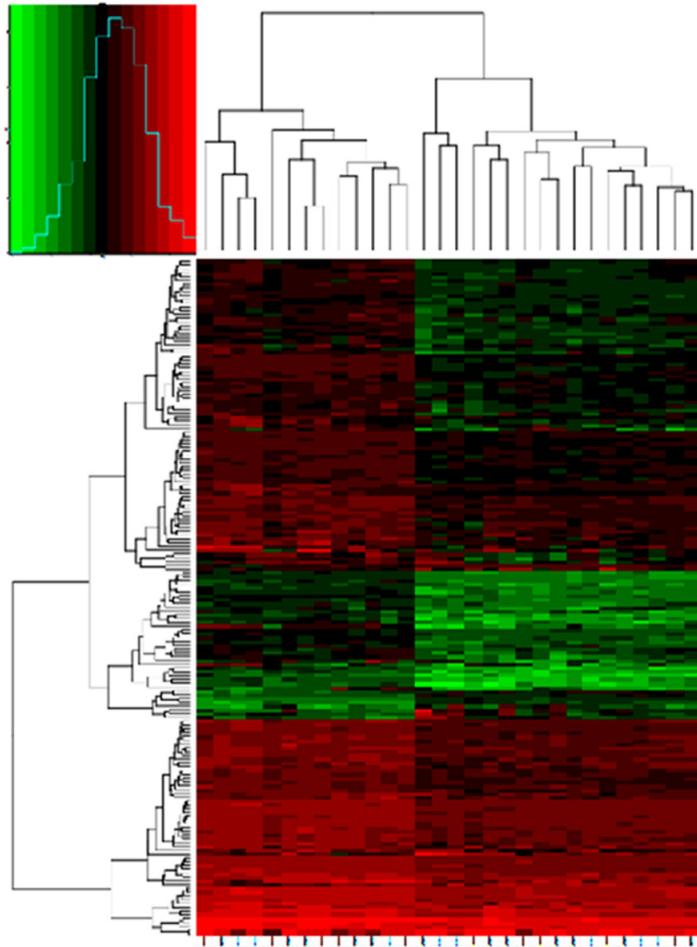
Figure 2. Heat map of differentially expressed genes between lung cancer tissues and normal lung tissues (green for low expression and red for high expression).

MNAT1 expression in normal lung tissues and lung cancer tissues

In the Oncomine database, data on the gene expression of the MNAT1 gene in lung cancer tissues and normal tissues were analysed, and the expression of this gene was significantly increased in lung cancer, in which it promotes the development of cancer (**Figure 7**).

High MNAT1 expression predicts a poor prognosis in patients with lung cancer

Data on the survival time, survival status and gene expression of the MNAT1 gene in lung cancer patients from the Oncomine database were downloaded. By analysing the expression level of the MNAT1 gene, lung cancer patients were divided into a high expression group and a low expression group. The survival rate was lower in the high expression group, whereas it was higher in the low expression group (**Figure 8**).

Discussion

With the ageing of the global population, the number of cancer patients is gradually increasing. Global attention to cancer patients has been widely observed. Since 2015, the "Precision Medicine Initiative" proposed by US President Barack Obama has become the focus of national research [11]. Committed to cancer patients and diabetic patients [12], China has also proposed a precise medical strategic plan [7]. High-throughput sequencing technology, omics technology [13], big data mining, and analysis technology play an important role in tumour research for the precise treatment of tumours [14]. This study mainly uses these related technologies to explore the role of the key gene MNAT1 in the development of lung cancer in tumours.

MNAT1 (ménage a trois 1, MAT1) was initially identified as the third subunit in addition to CDK7 and cyclin H in the cyclin-dependent kinase-activating kinase (CAK) complex [15]. MNAT1 exerts the above functions through its distinct domains interacting with downstream molecules. The C-terminal domain of MNAT1 interacts with the CDK7-cyclin H complex to stimulate CDK7 kinase activity. The coiled-coil domain of MNAT1 interacts with XPD and XPB to anchor CAK to the TFIIH core, while the N-terminal domain RING finger of MNAT1 is involved in C-terminal domain (CTD) phosphorylation of RNA Polymerase II (Pol II), which is required for gene promoter release and transcription initiation [16]. Intact MNAT1 expression is associated with cell cycle G1 exit, whereas intrinsically programmed or RA-induced MNAT1 degradation leads to cell cycle arrest, transcription inhibition, and cell differentiation [17]. Suppressed MNAT1 triggers apoptosis
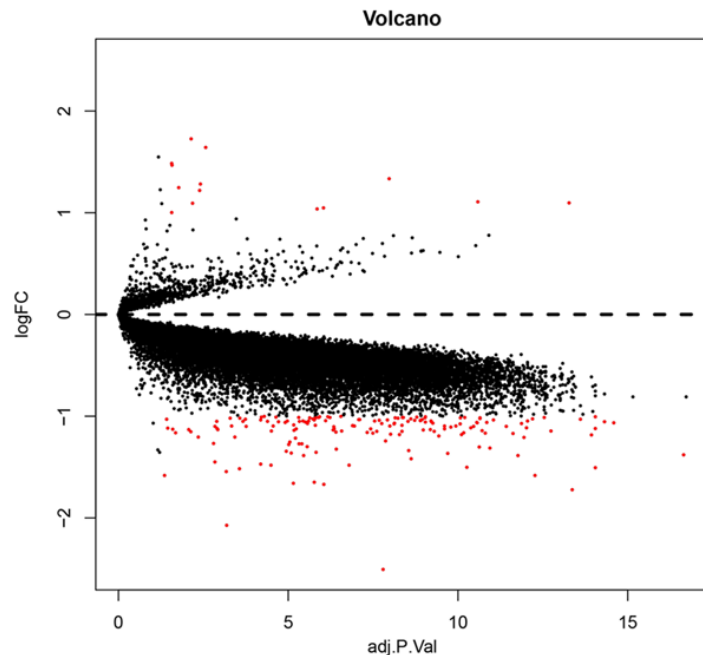
Figure 3. Volcano plot of differentially expressed genes between lung cancer tissues and normal lung tissues (black is no difference, red is differential).

[18]. Recent reports show that MNAT1 is over-expressed in breast cancer, and its expression level is associated with ER expression and patient outcome [19]. Shan Zhou found that MNAT1 highly regulated p53 downstream molecules, including p21, PARK, RAD51, and FAS, of which p21 is critical for p53-mediated G1/S boundary cell cycle arrest [20]. MNAT1 not only decreased BAX but also mediated the increase in FAS. FAS is a component transcriptionally regulated by p53 in the extrinsic apoptotic pathway [21]. MNAT1 interacts with p53 and promotes the ubiquitination and degradation of p53. MNAT1-mediated p53 degradation may be critical for CRC initiation and progression. The activity of p53 can be regulated through ubiquitination, oxidation, phosphorylation, acetylation and methylation [22, 23].

In this study, data of the genes from lung cancer patients and non-lung cancer patients were downloaded from the GEO data platform and analysed. The differentially expressed genes between the two groups were analysed, and finally, the two groups were selected. A total of 178 differentially expressed genes were found, 14 of which had an increased expression (LOC102723448, ITGA9-AS1, NUPR1, IFI35, IFI35, KIRREL3, MNAT1, FOLR3, FOLR3, CA3,

MMP12, SCG2, SCG2, and ALDH1A2). There were 164 differentially expressed genes that had a low expression, including SLC-25A1, PPP1R15B, LOC1005072-91, WBSCR27, LIPT1, CLDN10, NEURL4, etc. GO analysis revealed that the differentially expressed genes in the two groups were mainly located in the nucleus, cytoplasm, and RNA polymerase regulating cells and had functions of transcriptional extension of the II promoter, gene excision repair, DNA repair, negative regulation of hepatocyte proliferation, transcription-coupled nucleic acid excision and repair. The study of Patel H [19] and other studies have shown that MNAT1, CDK7, and cyclin H can promote the development of breast cancer and play a negative role in the prognosis of breast cancer. Klupp F et al. [24] confirmed the prognosis of colon cancer patients with a high expression of MMP12 in serum, mainly by observing the invasive ability of the gene in the blood vessels of patients coupled with the five-year survival rate of patients, confirming its growth promotion to colon cancer. Therefore, MNAT1 and MMP12 [25] might also promote the development of cancer and affect its prognosis in lung cancer. Oncomine database was used to analyse the gene expression of 17 patients with non-lung cancer and 21 patients with lung cancer. Expression of this gene in lung cancer was higher than that in non-lung cancer patients, and there was a significant difference between the two groups (P < 0.05). At the same time, the survival time of lung cancer patients in this database was analyzed. The study included 124 patients with lung cancer, including 80 patients with a high expression of the MNAT1 gene, 44 patients with a low expression of this gene, and 72 patients in both groups who died. Among the patients who died, 47 were in the high expression group, and 25 were in the low expression group (**Figure 8**). The analysis showed that a high expression of the MNAT1 gene in lung cancer patients inhibited the prognosis of lung cancer patients.

To further investigate the mechanism by which the MNAT1 gene promotes the development of

Table 1. GO enrichment analysis of differential genes in lung cancer patients

Term	Count	Genes	FDR
dephosphorylation	4	DUSP28, NT5C3B, CA3, PON3	30.98
transcription elongation from RNA polymerase II promoter	4	MNAT1, ELL3, SUPT5H, GTF2H1	30.98
DNA repair	6	MNAT1, EID3, LIG1, CINP, FANCG, GTF2H1	6
identical protein binding	11	SYP, S100A4, RALYL, PCDHB8, EID3, CLDN10, NDC80, PTS, NQO1, CLDN14, TRIM21	49.13
cell cycle	5	MNAT1, CINP, SUPT5H, MCTS1, TRIM21	67.65
nucleoplasm	27	GPN1, EID3, FERMT2, SHOC2, PEX19, ISG15, CASP7, PSMB3, DNAJC8, FBXO5, RGN, FANCG, SUPT5H, NTHL1, RTCA, NEIL3, LIG1, FDPS, CELSR1, ELL3, TATDN1, GTF2H1, SLTM, MNAT1, DCLRE1A, DACT1, HDAC1	66.71
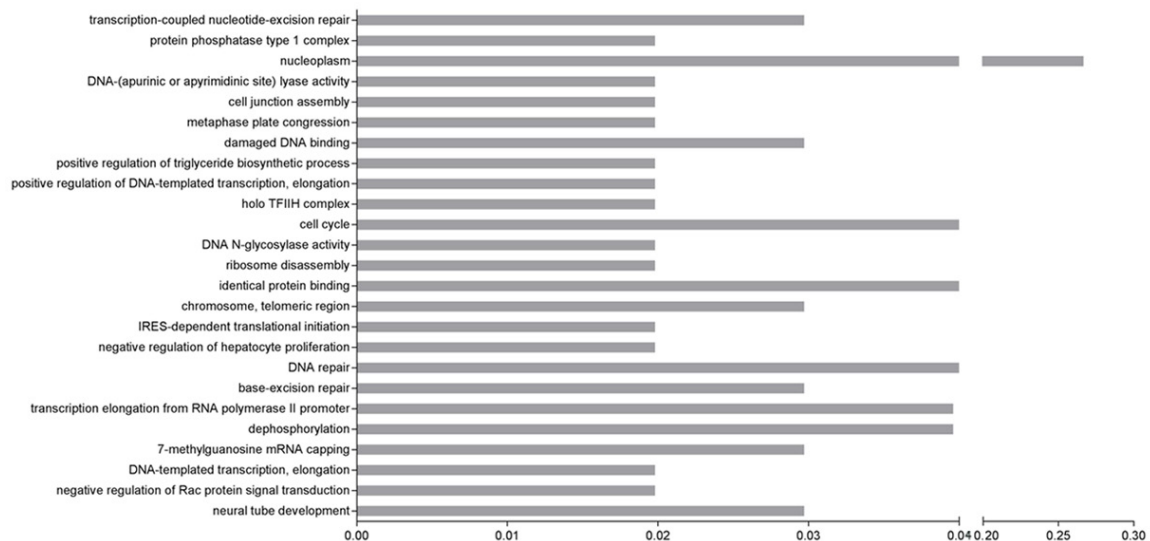
Figure 4. GO enrichment analysis of differentially expressed genes from the GSE54495 dataset.

Table 2. Analysis of KEGG pathway of differential genes in patients with non-small cell lung cancer

Term	Count	Genes	FDR
Complement and coagulation cascades	4	F5, C3, C4BPA, CPB2	15.85
Base excision repair	3	LIG1, NEIL3, NTHL1	25.21
Nucleotide excision repair	3	MNAT1, LIG1, GTF2H1	43.04
Cell adhesion molecules (CAMs)	4	LRRC4, CD8A, CLDN10, CLDN14	66.39
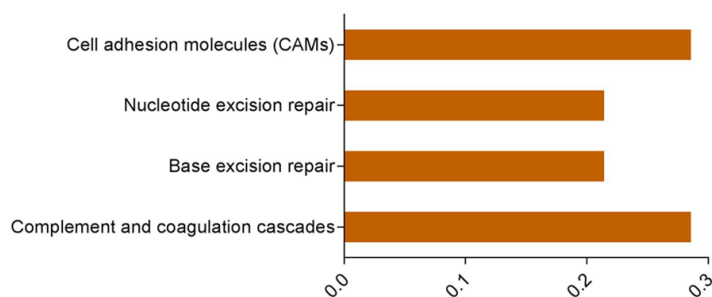
Figure 5. KEGG signalling pathway analysis of differentially expressed genes in the GSE54495 dataset.

lung cancer, the signalling pathway of this gene was analyzed and found to be mainly related to the nucleotide excision repair (NER) pathway. The nucleotide excision repair pathway is the most important pathway in the DNA damage repair pathway [26]. Its main function is to remove DNA adducts caused by ultraviolet radiation or other physical and chemical carcinogens in the environment and

Figure 6. Schematic diagram of differential protein interaction.

Figure 7. Differential expression patterns of MNAT1 in lung cancer tissues and normal lung tissues.

to repair damaged DNA regions. MNAT1, due to its variety of repairable DNA damage, has a wide range of functions and is being noticed by more and more researchers [27].

In summary, this study used bioinformatics methods to analyse the high expression of the MNAT1 gene in promoting the development of lung cancer and inhibiting the survival rate of patients by acting on the DNA repair mechanism. However, this study lacks corresponding experimental confirmation. Additional confirmatory studies are planned to define the mechanism of the MNAT1 gene in lung cancer through experimental methods in subsequent research and further fully explain the important role of bioinformatics in speculating oncogenes. These methods complement each other for better general and clinical applications.
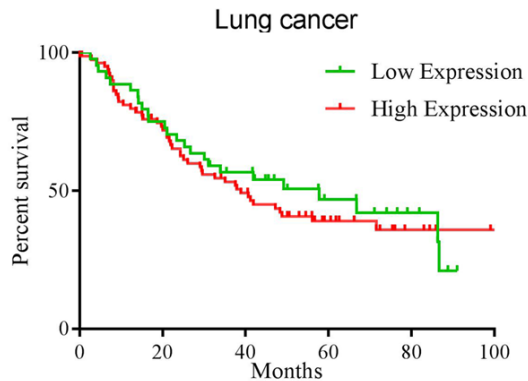
Figure 8. Kaplan-Meier analysis of the relationship between MNAT1 expression and the survival rates of lung cancer patients with different MNAT1 expression patterns (green represents patients with a low expression of MNAT1, red represents patients with a high expression of MNAT1).

Disclosure of conflict of interest

None.

Address correspondence to: Gang Chen, Department of Respiration, The Third Hospital of Hebei Medical University, Shijiazhaung 050051, China. E-mail: chengang_hb@163.com; Lifei Wang, Department of Respiration, The Third Hospital of Changzhou, Changzhou 213003, China. E-mail: 1063868887@qq.com

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018; 68: 7-30.

[2] Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med 2011; 32: 605-644.

[3] Li J, Shi R, Yu P. Effects of chemotherapy on global gene expression in non-small cell lung cancer. Eur Rev Med Pharmacol Sci 2014; 18: 126-131.

[4] Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M1, Lancet D. The genecards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016; 54: 1.

[5] Gunderson CC, Rowland MR, Wright DL, Andrade KL, Mannel RS, Mcmeekin DS and Moore KN. Initiation of a formalized precision medicine program in gynecologic oncology. Gynecol Oncol 2016; 141: 24-28.

[6] Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015; 372: 793-795.

[7] Sara R. Precision-medicine plan raises hopes. Nature 2015; 517: 540.

[8] Fan W, Pritchard JI, Olson JM, Khalid N and Zhao LP. A class of models for analyzing GeneChip gene expression analysis array data. BMC Genomics 2005; 6: 16.

[9] Larsson O, Wahlestedt C and Timmons JA. Considerations when using the significance analysis of microarrays (SAM) algorithm. BMC Bioinformatics 2005; 6: 129.

[10] Zhang J, Li C, Wu C, Xiong L, Chen G, Zhang Q and Wang S. RMD: a rice mutant database for functional analysis of the rice genome. Nucleic Acids Res 2006; 34: D745-748.

[11] Terry SF. Obama's precision medicine initiative. Genet Test Mol Biomarkers 2015; 19: 113-114.

[12] McCarthy M. Obama promises to defend health law and promote "precision medicine". BMJ 2015; 350: h385.

[13] Zhu QL, Shi L, Peng G and Luan FS. High-throughput sequencing technology and its application. Journal of Northeast Agricultural University 2014; 21: 84-96.

[14] Latkowski T and Osowski S. Data mining for feature selection in gene expression autism data. Expert Systems with Applications 2015; 42: 864-872.

[15] Fujii W, Nishimura T, Kano K, Sugiura K and Naito K. CDK7 and CCNH are components of CDK-activating kinase and are required for meiotic progression of pig oocytes. Biol Reprod 2011; 85: 1124-1132.

[16] Busso D, Keriel A, Sandrock B, Poterszman A, Gileadi O, Egly JM. Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. J Biol Chem 2000; 275: 22815-22823.

[17] Wu L, Chen P, Shum CH, Chen C, Barsky LW, Weinberg KI, Jong A, Triche TJ. MAT1-modulated CAK activity regulates cell cycle G(1) exit. Mol Cell Biol 2001; 21: 260-270.

[18] Wu L, Chen P, Hwang JJ, Barsky LW, Weinberg KI, Jong A, Starnes VA. RNA antisense abrogation of MAT1 induces G1 phase arrest and triggers apoptosis in aortic smooth muscle cells. J Biol Chem 1999; 274: 5564-5572.

[19] Patel H, Abduljabbar R, Lai CF, Periyasamy M, Harrod A, Gemma C, Steel JH, Patel N, Busonero C, Jerjees D, Remenyi J, Smith S, Gomm JJ, Magnani L, Győrffy B, Jones LJ, Fuller-Pace F, Shousha S, Buluwela L, Rakha EA, Ellis IO, Coombes RC, Ali S. Expression of CDK7, Cyclin H, and MAT1 is elevated in breast cancer and Is prognostic in estrogen receptor-positive breast cancer. Clin Cancer Res 2016; 22: 5929-5938.

[20] Abbas T, Dutta A. p21 in cancer: intricate networks and multiple activities. Nat Rev Cancer 2009; 9: 400-414.

[21] Aubrey BJ, Kelly GL, Janic A, Herold MJ and Strasser A. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? Cell Death Differ 2018; 25: 104-113

[22] Brooks CL and Gu W. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol 2003; 15: 164-171.

[23] Ranjan K and Pathak C. Expression of cFLIP L determines the basal interaction of Bcl-2 with Beclin-1 and regulates p53 dependent ubiquitination of Beclin-1 during autophagic stress. J Cell Biochem 2016; 117: 1757-1768.

[24] Klupp F, Neumann L, Kahlert C, Diers J, Halama N, Franz C, Schmidt T, Koch M, Weitz J, Schneider M, Ulrich A. Serum MMP7, MMP10 and MMP12 level as negative prognostic markers in colon cancer patients. BMC Cancer 2016; 16: 494.

[25] Grudny J, Kołakowski J, Kruszewski M, Szopiński J, Sliwiński P, Wiatr E, Winek J, Załęska J, Zych J, Roszkowski-Śliż K. Association of genetic dependences between lung cancer and chronic obstructive pulmonary disease. Pneumonol Alergol Pol 2013; 81: 308-318.

[26] Muñoz-Galván S, López-Saavedra A, Jackson SP, Huertas P, Cortés-Ledesma F, Aguilera A. Competing roles of DNA end resection and non-homologous end joining functions in the repair of replication-born double-strand breaks by sister-chromatid recombination. Nucleic Acids Res 2013; 41: 1669-1683.

[27] Zhi EL, Liang GQ, Li P, Chen HX, Tian RH, Xu P and Li Z. Seminal plasma miR-192a: a biomarker predicting successful resolution of nonobstructive azoospermia following varicocele repair. Asian J Androl 2018; 20: 396-399.