MTOR hypermethylation may associate with the susceptibility and survival of SARS-CoV-2 infections to lung adenocarcinoma patients based on multi-omics data and machine learning

2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-6819-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIBM55620.2022.9995080

Yu Guo[#]
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
guoyu@hit.edu.cn

Minghao Li[#]
Beidahuang Industry Group General Hospital
Harbin, China
liminghao0906@163.com

Yang Hu[*]
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
huyang@hit.edu.cn

Tianyi Zang[*]
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
tianyi.zang@hit.edu.cn

***Abstract*—Recent studies have shown that lung adenocarcinoma (LUAD) patients have a higher risk and worse prognosis of COVID-19 caused by SARS-CoV-2 compared to normal samples. Whereas, in addition to the receptor for SARS-CoV-2, other genes also deserve attention. In our study, we identified 19 differentially methylated genes (DMGs) that were co-upregulated in LUAD and COVID-19 samples. These 19 DMGs mainly regulated the immune-related and multiple viral infection signaling pathways. Gene Ontology and pathway enrichment analysis were applied with these genes. Then, 6 key DMGs (MTOR, ACE, IGF1, PTPRC, C3, and PTGS2) were identified by constructing and analyzing the protein-protein interaction (PPI) network. Besides, MTOR was significantly associated with 5 prognostic markers (CDO1, NEURL4, SMAP1, NPEPPS, IQCK) identified by survival analysis based on machine learning. In total, MTOR hypermethylation may be related to the susceptibility of LUAD patients to SARS-CoV-2 and the prognosis of LUAD patients suffering from COVID-19.**

***Keywords*—MTOR, LUAD, SARS-CoV-2, multi-omics, machine learning**

I. Introduction

Since December 2019, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has broken out worldwide [1]. SARS-CoV-2 is a highly infectious and pathogenic coronavirus. There were 616,427,419 people suffering from COVID-19, as of 5 October 2022, reported as WHO (https://covid19.who.int/). Patients with malignant tumors have low immune function and suppression of the systemic immune system caused by anti-tumor treatments such as radiotherapy and chemotherapy or surgery, which makes tumor patients more susceptible to SARS-CoV-2 than non-tumor patients [2], especially Lung cancer [3]. Among COVID-19 patients, lung cancer patients are the most common type of cancer [4]. Lung cancer is the cancer with the highest mortality rate in the world, among which more than 40% of cases are lung adenocarcinoma (LUAD) [5]. Additionally, COVID-19 patients with cancer are more likely to have acute complications than COVID-19 without cancer [6]. The underlying lung and immune dysfunction of LUAD patients will lead to a worse prognosis and higher mortality after being infected with SARS-CoV-2 [7]. So, it is significant to investigate the factors associated with the susceptibility and prognosis affecting SARS-CoV-2 infection in LUAD patients at molecular levels.

In our study, we applied multi-omics data of LUAD and COVID-19 downloaded from public database to discover the influence of MTOR methylation level on the survival and susceptibility of LUAD patients to SARS-Cov-2 infection. After preprocessing COVID-19 and LUAD DNA methylation datasets, we identified the co-upregulated or the co-downregulated common DMGs corresponding to DMPs. Based on these DMGs for further analysis, we applied Gene Ontology (GO) and KEGG pathway enrichment analysis to have an understanding of biological processes and functions. In order to detect key common genes, we utilized the construction and analysis of PPI network to extract hub genes and explore modules from common DMGs. Further, to explore the susceptibility factors of LUAD patients, we assessed the expression of MTOR in multi-omics data. Finally, we detected prognostic markers by survival analysis using machine learning.

II. Method

A. Data collection

Covid-19 dataset (GSE174818) illustrates infections of SARS-CoV-2 in DNA methylation level. It was downloaded from Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). The DNA methylation data and RNA-seq data of Lung adenocarcinoma (LUAD) were downloaded from The

© IEEE 2023. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

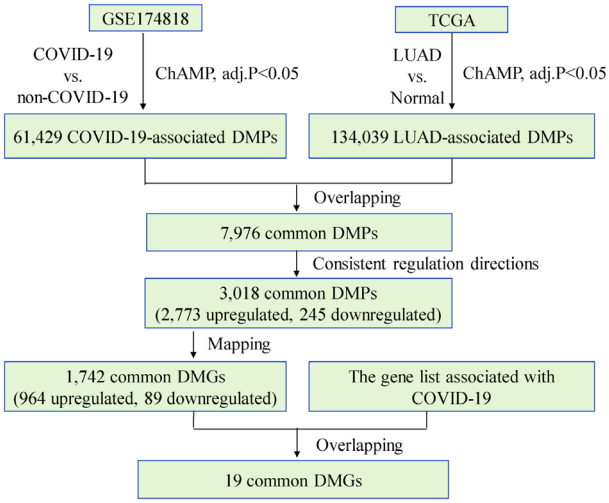

Fig. 1. Flowchart of LUAD and COVID-19 data processing.

Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/). The RNA-seq data of covid-19 were obtained from GEO, whose accession number is GSE152641. Illumina human MethylationEPIC data preprocessing

Illumina MethylationEPRIC microarray platform covers over 850,000 methylation sites. We downloaded raw.idat data from GEO. First, removed the probes with detection p value greater than 0.01. Second, filtered out probes with < 3 beads in at least 5% of samples per probe. Finally, filtered out all non-CpG probes, SNP-related probes, multi-hit probes and the probes located in sex chromosomes. After filtering probes, SWAN method was used to normalize the data. R package ChAMP was used to preprocess methylation microarray data in R version 4.1.1.

B. Identification of specific and common DMGs between LUAD and COVID-19 samples

Identification of DMGs in LUAD DNA methylation dataset and GSE174818 was the primary task of the research. To identify DMPs in both datasets, ChAMP package of R programming language was implemented. Cutoff criteria (P-value < 0.05) were applied to detect significant DMPs in these two datasets. We regarded a gene as DMG when there are at least one DMPs on this gene. The β-value of a gene is the mean value of the probes on the gene. In order to get genes most related to COVID-19, we downloaded the gene list which is the top 200 most searched genes associated with COVID-19 from DisGeNET (https://www.disgenet.org/home/) [8]. Furthermore, we regarded the intersect of the gene list and the DMGs as the COVID-19-associated genes. LUAD-associated genes were the DMGs in TCGA LUAD DNA methylation dataset. The common DMGs are the co-upregulated or co-downregulated DMGs between LUAD and COVID-19.

C. Gene ontology and pathway enrichment analysis

The gene ontology (GO) project provides an ontology of defined terms that represent the properties of gene products. GO covers three domains: cellular component, molecular function, and biological process. GO and pathway enrichment studies were conducted using the common DMGs between LUAD and COVID-19 by EnrichR (https://maayanlab.cloud/Enrichr/), which is a comprehensive gene set enrichment webtool [9]. Adjusted P-value < 0.05 was considered as a standard metric for quantifying the top ten listed Go terms and pathways, respectively.

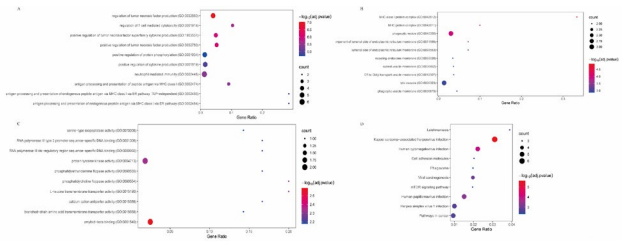

Fig. 2. Go and pathway enrichment analysis of 19 common DMGs. The top ten significant terms of (A) GO BP. (B) GO CC. (C) GO MF. (D) KEGG pathway.

D. Network construction and analysis

Proteins perform various important biological functions in organisms, but proteins usually do not function alone. They always function as team members in a dynamic network. There is growing evidence that protein-protein interactions are critical in many biological processes in cells. Protein-protein interaction (PPI) network of common DMGs was constructed by NetworkAnalyst (https://www.networkanalyst.ca/) [10], a visual analytics platform for comprehensive gene expression profiling and meta-analysis results, based on STRING database [11]. Next, the generated file was reintroduced into Cytoscape (version 3.8.2) to visualize and analyze the PPI networks. For PPIs network analysis, Hub genes in PPI network were detected by CytoHubba plugin in Cytoscape. Besides, highly dense modules were designed from the PPI network using MCODE plugin in Cytoscape.

E. DNA methylation prognostic analysis for LUAD

We downloaded the clinical information from TCGA. Then, we randomly split the 410 LUAD patents with complete survival information into training and validation data sets with an allocation of 3:1, corresponding to 274 and 136, respectively. Next, in the training dataset, we applied Univariate Cox and LASSO-Cox to screen markers for predicting survival outcome. A gene with adjusted P-value < 0.05 from Univariate Cox was retained in the dataset. Then, we used LASSO-Cox method to shrink the marker numbers to a reasonable range. Lambda was set as 0.1033725. The above analysis generated 5 final markers to construct a prognostic signature. By fitting a multi-variable Cox proportional hazards model on these 5 markers, we determined the coefficients of each marker and obtained a prognostic score for each individual. To validate our predictive model, we calculated the prognostic score for each individual in the validation dataset using coefficient estimates from the training dataset. By dividing the prognostic score according to its median, we formed high and low prognostic score groups with a roughly equal number of observations. We investigated if the median

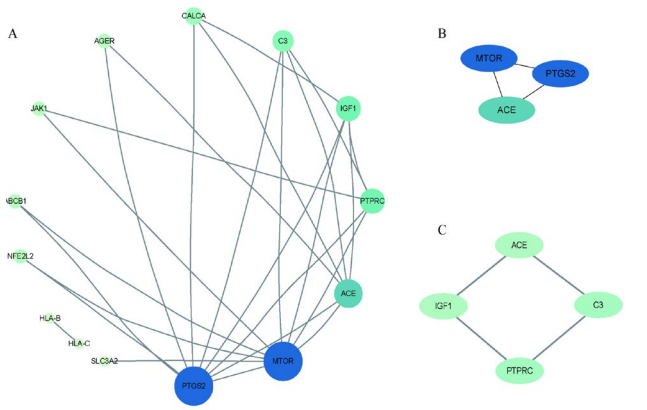

Fig. 3. The visualization of PPI network and its analysis results. (A) PPI network for 19 common DMGs. (B) 3 hub genes of the PPI network. (C) The module of the PPI network.

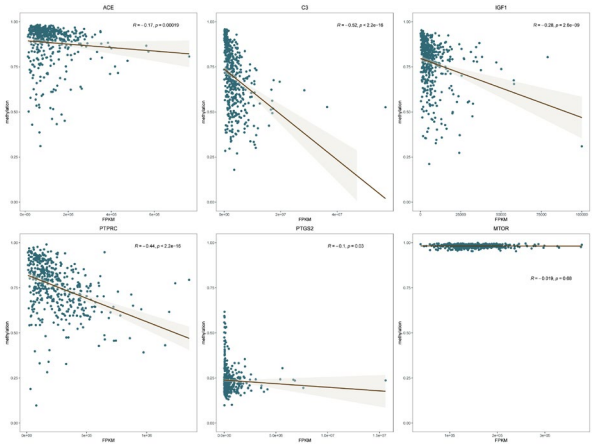

Fig. 4. The correlation between gene expression and the methylation level of the six genes associated with LUAD and SARS-CoV-2. (A) ACE (B) C3 (C) IGF1 (D) PTPRC. (E) PTGS2 (F) MTOR

survival time was significantly different between these two groups using a Kaplan-Meier estimator and log-rank test.

The analysis above was performed by the following R packages: glmnet, survival, survminer, ggplot2.

III. Results

A. Identification of specific and common DMGs between LUAD and COVID-19 patients

After preprocessing procedure, we identified 61,429 COVID-19-associated DMPs and 134,039 LUAD-associated DMPs, respectively, in GSE174818 and LUAD DNA methylation dataset (Fig. 1). In total, there are 7,976 common DMPs, of which 2,773 co-upregulated and 245 co-downregulated common DMPs. These 3,018 common DMPs mapped to 1,742 genes. Further, 964 co-upregulated DMGs and 89 co-downregulated DMGs were obtained. After intersecting with the gene list related to COVID-19, a total of 19 DMGs were obtained. These 19 DMGs were all co-upregulated significantly in both LUAD and COVID-19 patients, which may explain why LUAD patients are susceptible to SARS-CoV-2. This gene set of 19 common DMGs (*AGER*, *MTOR*, *NEU1*, *EPHB2*, *TWIST1*, *SLC3A2*, *ABCB1*, *IGF1*, *ACE*, *NFE2L2*, *SH2D3C*, *CALCA*, *KRT12*, *HLA-C*, *C3*, *PTPRC*, *HLA-B*, *PTGS2*, *JAK1*) was employed to accomplish further experiments.

B. Gene ontology and pathway enrichment analysis

The 19 common DMGs were used as the input gene set for Gene ontology and pathway enrichment analysis. GO analysis was acquired within three categories: biological process (BP), cellular component (CC), and molecular function (MF). The ongoing study illustrates the top 10 terms of Go for each of the subsections and pathway, which is depicted in Fig. 2.

C. PPI network to identify hub genes and module analysis

Recently, the construction of PPI networks has become an essential method in systems biology research. The PPI network (Fig. 3A) contains 14 nodes and 26 edges, of which 5 genes were deleted as the single nodes. The hub genes were sorted by their degree value, which indicates the number of interactions of the genes in the PPI network. Fig. 3B shows the top 3 identified hub genes: *MTOR*, *ACE* and *PTGS2*. The hub genes are important to the stability of the biological system and form a central part of the PPI network. Likewise, the module consists of 4 genes (*ACE*, *IGF1*, *PTPRC*, and *C3*) was depicted in Fig. 3C. Almost module genes showed prominent positions in the network, implying that these genes may play critical and similar roles. The hub gene *MTOR* is the gene encodes the protein belongs to a family of phosphatidylinositol kinase-related kinases. There are two primary mTOR inhibitors used in the treatment of human cancers, temsirolimus and everolimus. Besides, some studies have shown that mTOR inhibition might be the therapy against pandemic COVID-19 [12, 13].

D. The correlation between DNA methylation and gene expression in LUAD and COVID-19 samples

There is a large amount of evidence that DNA methylation plays a role in gene regulation. Importantly, DNA methylation in different genomic regions may have different effects on gene activities and it may influence the expression of genes and proteins. Here, we conduct further analysis on the six key genes (*MTOR*, *ACE*, *IGF1*, *PTPRC*, *PTGS2*, and *C3*) to detect the relationship of gene expression and DNA methylation. The DNA methylation levels of all key genes are negatively correlated with gene expression levels in LUAD samples, except for *MTOR* (Fig. 4).

E. Survival analysis for LUAD-associated DNA methylation markers

To explore whether the markers associated with LUAD likely to be of prognostic value, the predictive effect of each marker regarding overall survival (OS) for LUAD was subsequently assessed by performing survival analysis. The DNA methylation and clinical data of LUAD were downloaded from TCGA. Notably, we identified 5 prognosis methylation markers in LUAD using Univariate Cox and LASSO-Cox. We randomly assigned 410 LUAD samples to a training set (n = 274) and a testing set (n = 136) with the ratio 3:1. We implemented Univariate Cox and LASSO-Cox methods to reduce the dimensionality and

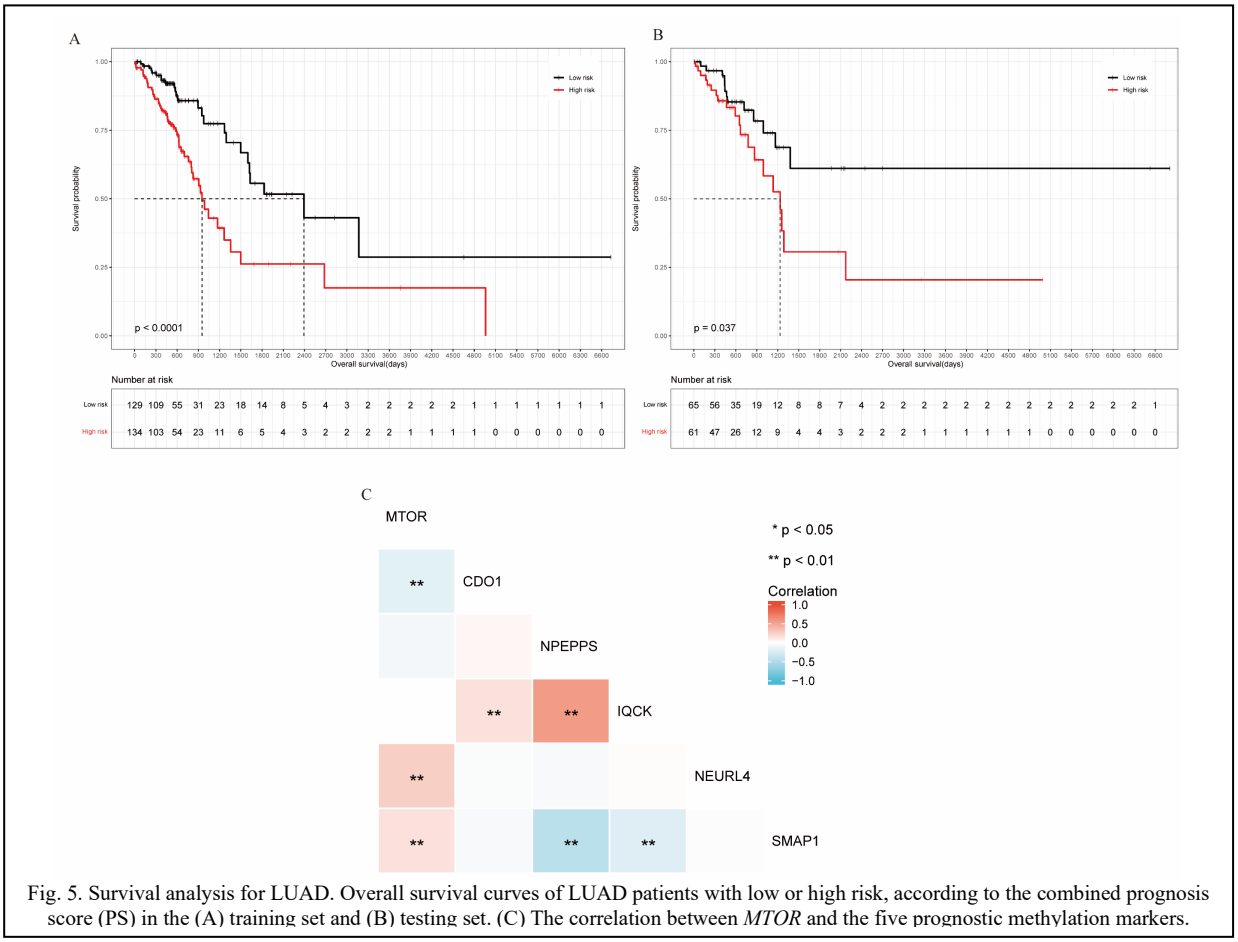

Fig. 5. Survival analysis for LUAD. Overall survival curves of LUAD patients with low or high risk, according to the combined prognosis score (PS) in the (A) training set and (B) testing set. (C) The correlation between *MTOR* and the five prognostic methylation markers.

constructed a Cox-model to predict prognosis with 5 markers. We generated Kaplan-Meier curves in training and validation data sets using a combined prognosis score, it shows as follow:

$$PS = 0.8916463 * CDO1 - 0.9477901 * NPEPPS - 0.2625149 * IQCK - 1.0761822 * NEURL4 + 0.4096752 * SMAP1$$

The threshold value for the mean of PS was -1.006183. Patients in the training set were divided into high-risk and low-risk groups based on their PS values. Consequently, patients with high-risk PS values exhibited a poorer prognosis compared with those with low-risk PS values in both training set and testing set (Fig. 5A-B). The DNA methylation level of *MTOR* has a significant correlation with *CDO1*, *NEURL4*, *SMAP1* (Fig. 5C). It reveals that *MTOR* may be associated with the prognosis of LUAD patients.

IV. Discussion

The global outbreak of COVID-19 caused by SARS-CoV-2 has become the most serious health threat worldwide, and has brought an unprecedented burden to the global medical system. COVID-19 is extremely contagious and has placed an unprecedented burden on the global medical system [14]. On the one hand, Lung cancer patients infected with SARS-CoV-2 have a significant increase in mortality [15]. On the other hand, lung cancer patients are more prone to respiratory tract infections, and often use immunosuppressive therapy. Additionally, the immune environment is disordered in LUAD, which is one of the reasons why LUAD patients are susceptible to SARS-CoV-2 [16, 17]. Despite most of the current researches focus on ACE2 and *TMPRSS2*, other genes and pathways also worthy of attention and investigation.

In our study, we identified 19 co-upregulated DMGs in COVID-19 and LUAD dataset. These DMGs were mainly enriched in “antigen processing and presentation of peptide antigen via MHC class I”, “mTOR signaling pathway” and multiple viral infections. This result was consistent with existing researches mentioned above. It showed that the 19 co-upregulated DMGs we identified that were associated to both covid and LUAD were credible. Subsequently, 6 key co-upregulated DMGs (*MTOR*, *ACE*, *IGF1*, *PTPRC*, *C3*, and *PTGS2*) were identified by PPI network construction and analysis. These 6 key co-upregulated DMGs in both datasets may indicate why LUAD patients were susceptible to SARS-CoV-2.

The multi-omics analysis results showed that the methylation level of *MTOR* was not significantly associated with its mRNA expression. However, there was no proteomic data of samples with paired mRNA or methylation data. We cannot explore the relationship of *MTOR* in multi-omics expression. We only showed the protein levels of *MTOR* were significantly in LUAD and normal samples. The effect of *MTOR* on protein expression and the relationship of multi-omics data still need to be further explored in COVID-19 and LUAD patients [18-20].

In total, this study has some limitations. Firstly, for multi-omics analysis, multi-omics data of the same samples should be used, strictly. Secondly, despite *MTOR*

participates in PI3K/AKT/ mTOR pathway, we didn't investigate the upstream and downstream of *MTOR* in this critical pathway[12]. The influence of *MTOR* on the upstream and downstream of the pathway deserves further discussions. Thirdly, immune infiltration analysis based on DNA methylation levels has been widely developed. It may provide more interesting conclusions [21, 22]. Finally, the above results were based on bioinformatics analysis and systems biology, which may need further experimental verifications.

ACKNOWLEDGMENT

We would like to acknowledge TCGA, GEO, and GTEx for providing multi-omics data.

REFERENCES

[1] C. C. Lai, T. P. Shih, W. C. Ko, H. J. Tang, and P. R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int J Antimicrob Agents,* vol. 55, no. 3, p. 105924, Mar 2020, doi: 10.1016/j.ijantimicag.2020.105924.

[2] J. Shi *et al.*, "Somatic Genomics and Clinical Features of Lung Adenocarcinoma: A Retrospective Study," *PLoS Med,* vol. 13, no. 12, p. e1002162, Dec 2016, doi: 10.1371/journal.pmed.1002162.

[3] W. Liang *et al.*, "Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China," *Lancet Oncol,* vol. 21, no. 3, pp. 335-337, Mar 2020, doi: 10.1016/S1470-2045(20)30096-6.

[4] B. Hu, H. Guo, P. Zhou, and Z. L. Shi, "Characteristics of SARS-CoV-2 and COVID-19," *Nat Rev Microbiol,* vol. 19, no. 3, pp. 141-154, Mar 2021, doi: 10.1038/s41579-020-00459-7.

[5] J. Yu, W. Ouyang, M. L. K. Chua, and C. Xie, "SARS-CoV-2 Transmission in Patients With Cancer at a Tertiary Care Hospital in Wuhan, China," *JAMA Oncol,* vol. 6, no. 7, pp. 1108-1110, Jul 1 2020, doi: 10.1001/jamaoncol.2020.0980.

[6] M. N. Uddin, R. Akter, M. Li, and Z. Abdelrahman, "Expression of SARS-COV-2 cell receptor gene ACE2 is associated with immunosuppression and metabolic reprogramming in lung adenocarcinoma based on bioinformatics analyses of gene expression profiles," *Chem Biol Interact,* vol. 335, p. 109370, Feb 1 2021, doi: 10.1016/j.cbi.2021.109370.

[7] V. Mehta *et al.*, "Case Fatality Rate of Cancer Patients with COVID-19 in a New York Hospital System," *Cancer Discov,* vol. 10, no. 7, pp. 935-941, Jul 2020, doi: 10.1158/2159-8290.CD-20-0516.

[8] J. Pinero *et al.*, "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res,* vol. 45, no. D1, pp. D833-D839, Jan 4 2017, doi: 10.1093/nar/gkw943.

[9] M. V. Kuleshov *et al.*, "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update," *Nucleic Acids Res,* vol. 44, no. W1, pp. W90-7, Jul 8 2016, doi: 10.1093/nar/gkw377.

[10] G. Zhou, O. Soufan, J. Ewald, R. E. W. Hancock, N. Basu, and J. Xia, "NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis," *Nucleic Acids Res,* vol. 47, no. W1, pp. W234-W241, Jul 2 2019, doi: 10.1093/nar/gkz240.

[11] D. Szklarczyk *et al.*, "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Res,* vol. 49, no. D1, pp. D605-D612, Jan 8 2021, doi: 10.1093/nar/gkaa1074.

[12] M. J. Ramaiah, "mTOR inhibition and p53 activation, microRNAs: The possible therapy against pandemic COVID-19," *Gene Rep,* vol. 20, p. 100765, Sep 2020, doi: 10.1016/j.genrep.2020.100765.

[13] B. S. Karam *et al.*, "mTOR inhibition in COVID-19: A commentary and review of efficacy in RNA viruses," *J Med Virol,* vol. 93, no. 4, pp. 1843-1846, Apr 2021, doi: 10.1002/jmv.26728.

[14] J. Sun *et al.*, "COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives," *Trends Mol Med,* vol. 26, no. 5, pp. 483-495, May 2020, doi: 10.1016/j.molmed.2020.02.008.

[15] J. Rogado *et al.*, "Covid-19 and lung cancer: A greater fatality rate?," *Lung Cancer,* vol. 146, pp. 19-22, Aug 2020, doi: 10.1016/j.lungcan.2020.05.034.

[16] A. Addeo, M. Obeid, and A. Friedlaender, "COVID-19 and lung cancer: risks, mechanisms and treatment interactions," *J Immunother Cancer,* vol. 8, no. 1, May 2020, doi: 10.1136/jitc-2020-000892.

[17] Q. Wang, N. A. Berger, and R. Xu, "Analyses of Risk, Racial Disparity, and Outcomes Among US Patients With Cancer and COVID-19 Infection," *JAMA Oncol,* vol. 7, no. 2, pp. 220-227, Feb 1 2021, doi: 10.1001/jamaoncol.2020.6178.

[18] H. Hua, Q. Kong, H. Zhang, J. Wang, T. Luo, and Y. Jiang, "Targeting mTOR for cancer therapy," *J Hematol Oncol,* vol. 12, no. 1, p. 71, Jul 5 2019, doi: 10.1186/s13045-019-0754-1.

[19] A. K. Murugan, "mTOR: Role in cancer, metastasis and drug resistance," *Semin Cancer Biol,* vol. 59, pp. 92-111, Dec 2019, doi: 10.1016/j.semcancer.2019.07.003.

[20] D. Mossmann, S. Park, and M. N. Hall, "mTOR signalling and cellular metabolism are mutual determinants in cancer," *Nat Rev Cancer,* vol. 18, no. 12, pp. 744-757, Dec 2018, doi: 10.1038/s41568-018-0074-8.

[21] S. C. Zheng, C. E. Breeze, S. Beck, and A. E. Teschendorff, "Identification of differentially methylated cell types in epigenome-wide association studies," *Nat Methods,* vol. 15, no. 12, pp. 1059-1066, Dec 2018, doi: 10.1038/s41592-018-0213-x.

[22] S. Chlamydas, A. G. Papavassiliou, and C. Piperi, "Epigenetic mechanisms regulating COVID-19

infection," *Epigenetics,* vol. 16, no. 3, pp. 263-270, Mar 2021, doi: 10.1080/15592294.2020.1796896.