The 9th IEEE International Conference on E-Health and Bioengineering - EHB 2021
Grigore T. Popa University of Medicine and Pharmacy, Web Conference, Romania, November 18-19, 2021

Lung Cancer Variants in Romanian population – a Genome-Wide Association Study

***Abstract* — Worldwide, lung cancer is the deadliest malignant tumour, being associated with a major risk factor: smoking. The exposure to environmental risk factors does not explain completely the development of lung malignant tumours, raising questions about the influence of the genetic background and the interaction between genes and environment. The study aims to identify variants which are associated with lung cancer in the Romanian population. After an association analysis performed in R Studio using PLINK tool for the genomic data obtained from 1386 cases and 1437 control subjects, 69 single nucleotide polymorphisms having a p value below 10^{-5} were identified, 4 of which had a p value below 10^{-8}. None of the variants identified has been previously reported as connected to lung cancer. After an investigation of the scientific literature regarding the genes where the statistically significant single nucleotide polymorphisms are located, multiple connections to lung cancer were identified.**

Keywords—GWAS, lung cancer, SNP, Romanian population, genetic variant.

I. Introduction

Lung cancer is the second most frequent type of cancer worldwide (2,6 million cases) and also the cause of most frequent cancer-related death (1,76 million cases) [1]. In Romania, estimated lung cancer incidence in 2020 was 12,7%, ranking fourth after breast, prostate and colorectal cancer[2]. Environmental factors such as smoking do not account for all causing factors, guiding the understanding of this disease to the genome. The particular genetic characteristics can offer information on how to create better strategies for the detection, prevention and treatment of the disease [3]. The aim of this study is to identify sequence variants associated with lung cancer in a genetically unscreened Romanian population by performing a genome-wide association study (GWAS).

II. Database and methods

A. Database

The subjects included were male and female patients who suffered from lung cancer (cases 1386) or different medical conditions (1437 controls), including hospital admission data from 2014 to 2018, from various clinics of Bucharest. All subjects gave their informed consent for the genetic research. The study was approved by the Bioethical Committee of the Romanian College of Physicians and National Ethical Board of the Romanian Medical Doctors Association in Romania.

Whole blood samples and buccal swabs were used for the DNA extraction at deCODE Genetics (Reykjavik, Iceland) and genotyped using Infinium OmniExpress-24 bead chips (Illumina). 716,503 SNPs (Single Nucleotide Polymorphisms) were eventually genotyped from all the individuals of the study.

B. Data analysis

R Studio was the software used for data processing, statistical analysis, and data visualization. Quality Control (QC) procedures were first performed using PLINK tool for GWASs. Excluding the individuals with more than 10% missing genotype data was the initial step in the QC process. Out of 5435 people initially loaded from the pedigree files, 50 were excluded due to the first criterion and only 104390 variants were further analyzed. Among remaining phenotypes, 1386 were cases and 1437 were controls. The next step implied the extraction of SNPs with less than 90% genotyping rate, 83 variants being removed after this filter. 104307 variants were additionally filtered for a minor allele frequency below 0.05. 92342 variants and 5385 subjects passed the QC so far. The final filter removed the markers that had failed the Hardy-Weinberg Equilibrium (HWE) test at a 0.001 significance threshold. 445 variants were extracted after the HWE filter. 91897 variants and 5385 subjects passed the QC tests and were included in the association studies.

PLINK tool was used to perform a standard case/control association analysis. To determine the significance of the contingency previously done, Fisher's exact test for allelic association was used. The file generated after Fisher's test served as marker for data visualization methods: using qqman and ggplot2 packages in R console, the final results were visually represented on a Manhattan Plot.

IEEE 978-1-6654-4000-4/21/$31.00

III. RESULTS

After data quality control and association tests, 69 SNPs identified (rs10493170, rs12568614, rs10494126, rs4656525, rs659580, rs340835, rs12565154, rs11899294, rs10194115, rs17741574, rs7593976, rs10179812, rs4674353, rs918233, rs873977, rs6550987, rs994439, rs9861303, rs1464311, rs837662, rs1545991, rs17625986, rs17002240, rs7680300, rs17832354, rs9631757, rs7662749, rs10039856, rs3923448, rs6912306, rs9320250, rs2753147, rs512077, rs17151696, rs6974424, rs10255742, rs36077344, rs4738193, rs17506603, rs448428, rs16915833, rs10121342, rs1346942, rs3124016, rs17703528, rs1905248, rs12309411, rs685167, rs1289392, rs4445762, rs1028529, rs7159803, rs17111121, rs11852212, rs1465163, rs7178742, rs11070918, rs2304973, rs2079578, rs9902289, rs7237580, rs2285957, rs9304911, rs11259990, rs11084727, rs7251374, rs11083522, rs2298393, rs535985). with a p value below 10^{-5} were identified (Figure 1). Four variants located on the chromosomes 1, 2, 8, 13 had a p value below 10^{-8} (rs10493170, rs17741574, rs16915833, rs4445762). (Figure 1). 17 SNPs had an Odds Ratio (OR) above 1, being identified in more subjects belonging to the case group, in comparison to the control group, and 52 SNPs had an OR below 1 (were identified prevalently in the control group).

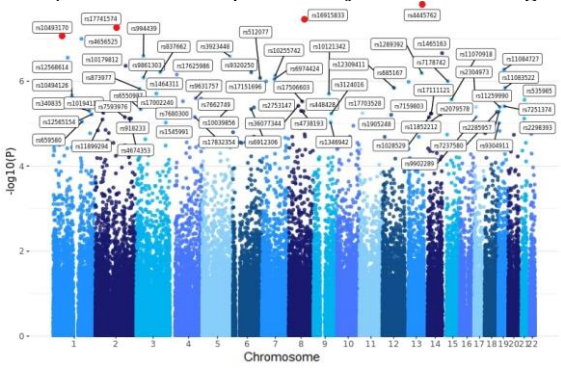

Fig. 1. Manhattan plot illustrating the statistically significant variants (the variants which had a p value below 10^{-8} are marked with red)

The statistically significant SNPs were investigated by accessing available databases: SNP database, ClinVar, GeneCards, GWAS Catalog. 52 SNPs were located in genes (39 encoding proteins and 13 encoding RNA species). 17 SNPs were located in the non-coding regions of the genome. The clinical relevance was reported in the used databases only for 3 SNPs, unconnected to lung cancer. This search did not identify previous reports regarding the implication in lung cancer of the 69 statistically significant variants. We further investigated the genes where these SNPs are located and multiple connections with lung cancer and carcinogenesis were discovered, providing evidence supporting our statistical results.

COL24A1 is a gene that is overexpressed in 10 tumours, including lung adenocarcinoma [5], [6]. We identified a statistically significant variant in this gene (rs12568614). rs10494126 is a SNP in DRAM2 gene, which has an important role in suppressing the expression of p53, thus functioning as an oncogene in non-small cell lung cancer. This gene is overexpressed in advanced stages of metastasis and it increases the production of factors involved in cell cycle and migration [7].

rs659580 is a variant in the PRRX1 gene. Its down-regulation may increase antiapoptotic properties of lung cancer cells, leading to cisplatin resistance. PRRX1 is connected to epithelial-mesenchymal transition (EMT), which is a process involved in tumour metastasis. The relationship between PRRX and EMT is controversial; some studies identified this gene as an inducer of EMT, while other studies proved that a loss of PRRX1 was associated with EMT, suggesting that the gene is an inhibitor of this process. PRRX1A isoform is upregulated in lung cancer tissue. It is also correlated with an increase in TGF beta expression [8]-[11].

rs340835 is a polymorphism located in the PROX1 gene, which is highly expressed in small cell lung cancer cell lines. A loss of this gene reduces the rate of these cells proliferation [12], [13]. rs918233 is a polymorphism of the CPS1 gene, which maintains DNA synthesis in KRAS/LKB-1-mutant lung cancer cells. This gene is upregulated in lung adenocarcinoma. CPS1 might be used as a potential therapy target and prognostic biomarker [14]-[17].

rs17625986 is a variant located in the RBM47 gene, which enhances the activity of p53 promoter, acting like a tumour suppressor. The gene can suppress the tumour growth by inhibiting Nrf2, being upregulated in non-small cell lung cancer [18]-[20]. rs9320250 is a SNP located in the OSTM1 gene, which has interactions with the RGS17 gene, a rare highly penetrant lung cancer gene [21], [22].

MAGI2-AS3 is another investigated gene link to rs10255742, a statistically significant SNP found in this gene. MAGI2-AS3 inhibits non-small cell lung cancer invasion and migration. This gene is downregulated in non-small cell lung cancer. An overexpression of MAGI2-AS3 can inhibit the proliferative and invasion properties of non-small cell lung cancer, promoting cell apoptosis [23]-[26]. rs36077344 is a variant of the PLAG1 gene, which acts as a transcription factor for GDH1, whose expression is correlated with the metastasis of lung cancer [27], [28].

MSC-AS1 is a gene involved in multiple types of cancer: lung adenocarcinoma, laryngeal cancer, pancreatic, gastric, kidney, nasopharyngeal, hepatocellular carcinoma and osteosarcoma. This gene is overexpressed in lung adenocarcinoma tissue and it facilitates the progression of lung adenocarcinoma [28]-[36]. We identified a statistically significant SNP of this gene (rs473819). Another SNP identified by our study, rs448428, is located in the ATP6V0D2 gene, whose expression in tumour associated macrophages in lung malignancies is suppressed by the lactate produced by cancer cells, resulting in the promotion of tumour vascularisation and growth [37].

rs685167 is a variant of the LATS2 gene, which functions as a tumour suppressor in non-small cell lung carcinoma. The gene is downregulated in lung cancer cell lines. LATS2 is directly inhibited by miR-25 in non-small

cell lung carcinoma, leading to increased cell proliferation, migration and invasion. When the gene is downregulated, it is associated with aggressive behaviour of the malignancy and worse prognosis [38]-[40]. rs2304973 is a polymorphism located in the CXCL16 gene, highly expressed in the lung cancer tissue, it facilitates proliferation and invasion of the malignant cells and it is correlated with poor prognosis of pulmonary cancer. CXCL16 expression is correlated with TNM stage [41]-[44]. MYOCD is a gene involved in promoting epithelial-mesenchymal transition that occurs in non-small cell lung cancer. rs9902289 is a variant of this gene identified by our study [45].

The expression of TIAM1 gene is higher in the lung cancer tissue than in normal tissues and it might be related to tumour development and cancer metastasis. TIAM1 promotes epithelial-mesenchymal transition and it increases angiogenesis, leading to lung adenocarcinoma progression. A statistically significant polymorphism (rs535985) was identified by our study. The overexpression of this gene is associated with worse prognosis [46], [47]. PEPD is a gene that suppresses p53. A loss of its activity is associated with cell death and regression of tumours. We identified a statistically significant variant of this gene (rs11084727) [48].

All the information presented above suggests that the results of this study are more than random associations of SNPs and lung cancer. Using the data regarding the functional consequences of the variants and the studies that connect some of the genes with lung cancer, we identified a few chains of biological processes that may explain the results of this study (Figure 2). The number of such connections we made was limited by the lack of data on the functional consequences of SNPs.

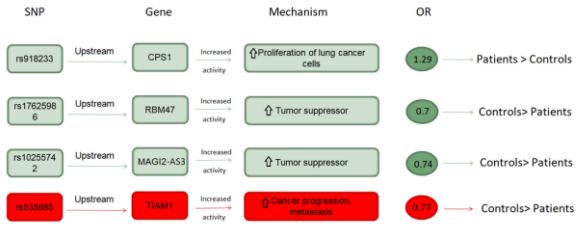

Fig. 2. Potential effects of 4 genetic variants on cancer development, in relationship with the statistical findings of the study

Rs918233 is an upstream variant that is associated with increased activity of CPS1, therefore this SNP can be related to a stimulation of the proliferation of lung cancer cells [14]-[17]. We obtained an OR of 1.29, indicating that it was observed more frequently in the lung cancer patients` group. Rs17625986 is an upstream polymorphism of RBM47 gene, its presence enhances the activity of this gene, and by this it can facilitate tumour suppression [18]-[20]. This can be a protective variant, this affirmation being supported by an OR of 0.7, indicating a higher prevalence of this variant in the control group. A similar mechanism can be described for the rs10255742 SNP of MAGI2-AS3 gene [23]-[26], which had an OR of 0.74. One biological mechanism we identified did not confirm our results: rs535985 is an upstream variant, which is associated with higher activity of the TIAM1 gene, which can lead to cancer progression [46], [47]. The OR we obtained was 0.77, indicating that this SNP was more prevalent in the control group, a result which is not consistent with the data obtained from previous research.

IV. Conclusions

Being one of the first GWAS targeted on Romanian subjects, this study provides a better understanding of the genetic component of lung cancer, specifically adjusted to the overall genetic background of the Romanian population. We have identified 69 SNPs with p value below 10^{-5}, 4 of them having a p value below 10^{-8}. We highlighted 16 of them as possible clinically relevant SNPs for lung cancer pathology, after investigating the corresponding genes, which are reported in literature as being connected to lung cancer. Further research should be conducted in order to determine the complex involvement of these polymorphisms in the development, survival, proliferation, and differentiation of malignant lung cells in cancer.

Acknowledgment

This study was funded by the EEA grant "Integrated Applied Genetics Training – AppGENEdu" Project, ID: EY-COP-0029, with data gathered in the EEA grant - ROMCAN project RO14-0017 EEAJRP-RO-NO-20131-10191.

References

[1] World Health Organisation, `Cancer`, Retrieved April 5th, 2021 from https://www.who.int/news-room/fact-sheets/detail/cancer

[2] European Cancer Information System, `Estimates of cancer incidence and mortality in 2020, for all countries`, Retrieved April 5th, 2021 from https://ecis.jrc.ec.europa.eu/explorer.php

[3] C.S. Dela Cruz, L.T. Tanoue, R.A. Matthay. `Lung cancer: epidemiology, etiology, and prevention`. Clinics in chest medicine, **32(4)**, 2011, pp. 605–644.

[4] B. Liang, H. Ding, L. Huang, H. Luo, X. Zhu. `GWAS in cancer: progress and challenges`. Mol Genet Genomics, **295(3)**, 2020, pp. 537-561.

[5] L. Yan, F. Xu, C. Dai. `Overexpression of COL24A1 in hepatocellular carcinoma predicts poor prognosis: a study based on multiple databases, clinical samples and cell lines`. OncoTargets and therapy, **13**, 2020, pp. 2819–2832.

[6] K. Misawa, T. Kanazawa, A. Imai, et al. `Prognostic value of type XXII and XXIV collagen mRNA expression in head and neck cancer patients`. Molecular and clinical oncology, **2(2)**, 2014, pp. 285–291.

[7] M. Wudu, H. Ren, L. Hui, et al. `DRAM2 acts as an oncogene in non-small cell lung cancer and suppresses the expression of p53`. Journal of experimental & clinical cancer research : CR, **38(1)**, 2019, pp. 72.

[8] L. Sun, T. Han, X. Zhang, et al. `PRRX1 isoform PRRX1A regulates the stemness phenotype and epithelial-mesenchymal transition (EMT) of cancer stem-like cells (CSCs) derived from non-small cell lung cancer (NSCLC)`. Translational lung cancer research, **9(3)**, 2020, pp. 731–744.

[9] H. Zhu, G. Sun, J. Dong, L. Fei. The role of PRRX1 in the apoptosis of A549 cells induced by cisplatin. American journal of translational research, **9(2)**, 2017, pp. 396–402.

[10] H. Zhu, G. Sun. `Loss of PRRX1 induces epithelial-mesenchymal transition and cancer stem cell-like properties in A549 cells`.

American journal of translational research, **9(4)**, 2017, pp. 1641–1650.

[11] Y. Tang, Y. Lu, Y. Chen, et al. `Pre-metastatic niche triggers SDF-1/CXCR4 axis and promotes organ colonisation by hepatocellular circulating tumour cells via downregulation of Prrx1`. Journal of experimental & clinical cancer research : CR, **38(1)**, 2019, 473.

[12] S.H. Zhu, C.J. Shan, Z.F. Whu, S.Z. Xu. `Proliferation of small cell lung cancer cell line reduced by knocking-down PROX1 via shRNA in Lentivirus`. Anticancer Research, **33(8)**, 2013, pp. 3169-3175;.

[13] C. Chang, P. Wang, H. Yang , L. Li, L.B. Zhang. `Expression of LYVE-1 and Prox-1 in non-small cell lung cancer and the relationship with lymph node metastasis`. Sichuan Da Xue Xue Bao Yi Xue Ban, **42(2)**, 2011, pp. 174-8.

[14] J. Kim, Z. Hu, L. Cai, et al. (2017). `CPS1 maintains pyrimidine pools and DNA synthesis in KRAS/LKB1-mutant lung cancer cells`. Nature, **546(7656)**, 2017, pp. 168–172.

[15] G. Wu, Z. Zhao, Y. Yan, et al. `CPS1 expression and its prognostic significance in lung adenocarcinoma`. Annals of translational medicine, **8(6)**, 2020, pp. 341.

[16] M. Çeliktas, I. Tanaka, S.C. Tripathi, et al. `Role of CPS1 in cell growth, metabolism and prognosis in LKB1-inactivated lung adenocarcinoma`. Journal of the National Cancer Institute, **109(3)**, 2017, pp. 1–9.

[17] J. Yue, Q. Dai, S. Hao, et al. `Suppression of the NTS-CPS1 regulatory axis by AFF1 in lung adenocarcinoma cells`. J Biol Chem, **296**, 2021, pp. 100319.

[18] R. Li, H. Li, C. Ge, et al. `Increased expression of the RNA-binding motif protein 47 predicts poor prognosis in non-small-cell lung cancer`. Oncology letters, **19(4)**, 2020, pp. 3111–3122.

[19] D.J. Shen, Y.H. Jiang, J.Q. Li, L.W. Xu, K.Y. Tao. `The RNA-binding protein RBM47 inhibits non-small cell lung carcinoma metastasis through modulation of AXIN1 mRNA stability and Wnt/β-catentin signaling`. Surg Oncol, **34**, 2020, pp. 31-39.

[20] T. Sakurai, K. Isogaya, S. Sakai, et al. `RNA-binding motif protein 47 inhibits Nrf2 activity to suppress tumor growth in lung adenocarcinoma`. Oncogene, **35(38)**, 2016, pp. 5000–5009.

[21] C.R. Bodle, D.I. Mackie, D.L. Roman. `RGS17: an emerging therapeutic target for lung and prostate cancers`. Future medicinal chemistry, 5(9), 2013, pp. 995–1007.

[22] GeneCards, `OSMT1 Gene`, Retrieved March, 2nd, 2021 from https://www.genecards.org/cgi-bin/carddisp.pl?gene=OSTM1

[23] Y. Sui, W. Chi, L. Feng, J. Jiang. `LncRNA MAGI2-AS3 is downregulated in non-small cell lung cancer and may be a sponge of miR-25`. BMC pulmonary medicine, **20(1)**, 2020, pp. 59.

[24] F. Li, Q. Hu, Z. Pang, X. Xu. `LncRNA MAGI2-AS3 upregulates cytokine signaling 1 by sponging miR-155 in non-small cell lung cancer`. Cancer Biother Radiopharm, **35(1)**, 2020, pp. 72-76.

[25] X.Z. Hao, K. Yang. `LncRNA MAGI2-AS3 suppresses the proliferation and invasion of non-small cell lung carcinoma through miRNA-23a-3p/PTEN axis`. Eur Rev Med Pharmacol Sci, **23(17)**, 2019, pp. 7399-7407.

[26] J. He, X. Zhou, L. Li, Z. Han. `Long noncoding MAGI2-AS3 suppresses several cellular processes of lung squamous cell carcinoma cells by regulating miR-374a/b-5p/CADM2 axis`. Cancer management and research, **12**, 2020, pp. 289–302.

[27] L. Jin, J. Chun, C. Pan, et al. `The PLAG1-GDH1 axis promotes anoikis resistance and tumor metastasis through CamKK2-AMPK signaling in LKB1-deficient lung cancer`. Molecular cell, **69(1)**, 2018, pp. 87–99.e7.

[28] Cancer Discovery editorial staff. `Glutaminolysis drives lung cancer metastasis via the PLAG1–GDH1 axis`, Cancer Discov, **8(2)**, 2018, pp. 135.

[29] S. Li, S. Yang, C. Qiu, D. Sun. `LncRNA MSC-AS1 facilitates lung adenocarcinoma through sponging miR-33b-5p to up-regulate GPAM`. Biochem Cell Biol, **99(2)**, 2021, pp. 241-248.

[30] Y. Liu, L. Li, X. Wu, et al. `MSC-AS1 induced cell growth and inflammatory mediators secretion through sponging miR-142-5p/DDX5 in gastric carcinoma`. Aging, **13(7)**, 2021, pp. 10387-10395.

[31] L. Zhang, G. Zhao, S. Ji, Q. Yuan, H. Zhou. `Downregulated long non-coding RNA MSC-AS1 inhibits osteosarcoma progression and increases sensitivity to cisplatin by binding to MicroRNA-142`. Med Sci Monit, **26**, 2020, pp. e921594.

[32] C. Cao, Q. Zhong, L. Lu, et al. Long noncoding RNA MSC-AS1 promotes hepatocellular carcinoma oncogenesis via inducing the expression of phosphoglycerate kinase 1. Cancer Med, **9(14)**, 2020, pp. 5174-5184.

[33] H. Yao, L. Yang, L. Tian, Y. Guo, Y. Li. `LncRNA MSC-AS1 aggravates nasopharyngeal carcinoma progression by targeting miR-524-5p/nuclear receptor subfamily 4 group A member 2 (NR4A2)`. Cancer Cell Int, **20**, 2020, pp. 138.

[34] Z. Hu, L. Li, P. Cheng, et al. `lncRNA MSC-AS1 activates Wnt/β-catenin signaling pathway to modulate cell proliferation and migration in kidney renal clear cell carcinoma via miR-3924/WNT5A`. J Cell Biochem, **121(10)**, 2020, pp. 4085-4093

[35] Y. Sun, P. Wang, W. Yang, Y. Shan, Q. Zhang, H. Wu. `The role of lncRNA MSC-AS1/miR-29b-3p axis-mediated CDK14 modulation in pancreatic cancer proliferation and Gemcitabine-induced apoptosis`. Cancer Biol Ther, **20(6)**, 2019, pp. 729-739.

[36] Y. Liu, W. Meng, H. Cao, B. Wang. `Identification of MSC-AS1, a novel lncRNA for the diagnosis of laryngeal cancer`. Eur Arch Otorhinolaryngol, **278(4)**, 2021, pp. 1107-1118.

[37] N. Liu, J. Luo, D. Kuang, et al. `Lactate inhibits ATP6V0d2 expression in tumor-associated macrophages to promote HIF-2α-mediated tumor progression`. The Journal of clinical investigation, **129(2)**, 2019, 631–646.

[38] Y. Xie, Y. Lv, Y. Zhang, Z. Liang, L. Han, Y. Xie.` LATS2 promotes apoptosis in non-small cell lung cancer A549 cells via triggering Mff-dependent mitochondrial fission and activating the JNK signaling pathway`. Biomed Pharmacother, **109**, 2019, pp. 679-689.

[39] T. Wu, H. Hu, T. Zhang, et al. `miR-25 promotes cell proliferation, migration, and invasion of non-small-cell lung cancer by targeting the LATS2/YAP signaling pathway`. Oxidative medicine and cellular longevity, 2019, pp. 9719723.

[40] S. Jang, M. Oh, H. Cho, et al. `Low LATS2 expression is associated with poor prognosis in non-small cell lung carcinoma`. Polish Journal of Pathology, **70(3)**, 2019, pp. 189-197.

[41] K. Liang, Y. Liu, D. Eer, J. Liu, F. Yang, K. Hu. `High CXC chemokine ligand 16 (CXCL16) expression promotes proliferation and metastasis of lung cancer via regulating the NF-κB pathway.` Medical science monitor : international medical journal of experimental and clinical research, **24**, 2018, pp. 405–411.

[42] Y. Shibata, N. Kobayashi, T. Sato, K. Nakashima, T. Kaneko. `The clinical significance of CXCL16 in the treatment of advanced non-small cell lung cancer`. Thorac Cancer, **11(5)**, 2020, pp. 1258-1264.

[43] C. Ke, Y. Ren, L. Lv, W. Hu, W. Zhou. `Association between CXCL16/CXCR6 expression and the clinicopathological features of patients with non-small cell lung cancer`. Oncology letters, **13(6)**, 2017, pp. 4661–4668.

[44] W. Hu, Y. Liu, W. Zhou, L. Si, L. Ren. `CXCL16 and CXCR6 are coexpressed in human lung cancer in vivo and mediate the invasion of lung cancer cell lines in vitro`. PloS one, **9(6)**, 2014, pp. e99056.

[45] X. Tong, S. Wang, Z. Lei, et al. `MYOCD and SMAD3/SMAD4 form a positive feedback loop and drive TGF-β-induced epithelial-mesenchymal transition in non-small cell lung cancer`. Oncogene, **39(14)**, 2020, pp. 2890-2904.

[46] H.M. Wang, J. Wang. `Expression of Tiam1 in lung cancer and its clinical significance`. Asian Pac J Cancer Prev, **13(2)**, 2012, pp. 613-5.

[47] G. Zhu, Y. Zhang, Q. Wang, et al. `The prognostic value of Tiam1 correlates with its roles in epithelial-mesenchymal transition progression and angiogenesis in lung adenocarcinoma`. Cancer management and research, **11**, 2019, pp. 1741–1752.

[48] L. Yang, Y. Li, A. Bhattacharya, Y. Zhang. `PEPD is a pivotal regulator of p53 tumor suppressor`. Nature communications, **8(1)**, 2017, pp. 2052.