Biotech Studies 30(1), 33-42
http://doi.org/xxx/xxx

PROOF
RESEARCH PAPER

BIOTECH STUDIES

Potential biomarkers for lung adenocarcinoma identified by integrative transcriptomics analysis

Ceyda Kasavi

Department of Bioengineering, Marmara University, 34722, Istanbul, Turkey.

Article History
Received 13 February 2021
Accepted 27 April 2021
First Online 03 May 2021

Corresponding Author
Tel.: +90 216 777 3566
E-mail:
ceyda.kasavi@marmara.edu.tr

Keywords
Lung adenocarcinoma
Transcriptome profiling
Network reconstruction
Potential biomarkers
Survival analysis

Abstract

Lung cancer is one of the most occurring and death-causing cancers worldwide. Despite the progress, survival rate is still low due to the late diagnosis. The aim of this study is to develop a computational framework to identify potential prognostic biomarkers for lung adenocarcinoma (LUAD). Gene expression profiles obtained from three independent studies were analyzed to find differentially expressed genes (DEGs) in LUAD. Disease-specific protein-protein interaction (PPI) network was constructed among common DEGs and hub proteins were identified. Gene expression data was integrated with the human transcriptional regulatory network (TRN) to identify key regulatory elements and construct disease-specific TRN. Hub proteins that were also present in TRN of LUAD were considered as potential biomarkers and assessed by survival analysis. *AURKA, CAV1, CLU, ENO1, FHL1, FHL2, LMO2, MYH11, NME1* and *SFN* were discovered as biomarkers for LUAD and survival analysis not only indicated their significant prognostic performance as a group but also revealed their contribution to the discrimination of risk groups. Our findings suggested that identified biomarkers could be valuable in LUAD progression and they should be considered for further experimentation.

Introduction

Lung cancer is one of the most malignant tumors and the leading cause of cancer-associated deaths (Yan et al., 2019). Lung cancer is mainly classified as small cell and non-small cell lung cancer (NSCLC). NSCLC, which accounts for approximately 85% of all the patients, has three main pathological subtypes, including adenocarcinoma (AD), squamous cell carcinoma (SCC), and large cell carcinoma (LCC). When compared to others, AD is the most common type (approximately 50%) (Cagle et al., 2013). Despite the increased understanding of the molecular mechanisms associated with lung cancer and the improvements in traditional treatment, the overall survival rate still remains low. Most patients are diagnosed in a late stage. Therefore, the early detection and diagnosis of the disease can improve the prognosis, increase the survival rate and lower the mortality rate (Villalobos & Wistuba, 2017; Yan et al., 2019).

Biological cancer markers are used for diagnostic, prognostic and treatment purposes (Villalobos & Wistuba, 2017). The targeted therapies in patients with *EGFR* mutations and/or *ALK* translocations improved survival (Patel et al., 2015), suggesting that the genomic biomarkers can serve not only for early diagnosis of the disease, but also for targeted therapies (Li et al., 2019; Villalobos & Wistuba, 2017). Moreover, the identification of the mutations, gene amplifications, deletions, or the presence of the fusion genes, which are accepted as the genetic risk factors associated with lung cancer would lead to developing treatments (Otálora-Otálora et al., 2019). Therefore, the search for novel biomarkers is of great significance and new methods to identify biomarkers with high prognostic performance need to be developed.

Microarray technology has been extensively used to identify genes with altered expressions during tumorigenesis, and therefore, enhances our understanding of the genetics and the molecular biology of cancer (Cieślik & Chinnaiyan, 2018). The identification of the transcriptionally dysregulated genes and their associated biological processes and signaling pathways allow us to understand the expression patterns related to tumor's biological state and the patients' survival (Otálora-Otálora et al., 2019). Moreover, the integration of transcriptome with biological and regulatory networks provide an opportunity to discover novel prognostic biomarkers and therapeutic strategies (Gov et al., 2017).

In the present study, a computational framework was developed to identify potential biomarkers for LUAD (Figure 1). For this purpose, firstly, gene expression data from three independent studies were analyzed and common DEGs, biological processes and molecular pathways involved in LUAD were determined. Secondly, a disease-specific PPI network was reconstructed with common DEGs and hub proteins were determined via simultaneous analysis of twelve node scoring metrics. Then, key regulatory elements involved in LUAD were identified by statistical analysis using hypergeometric probability distribution function and disease-specific TRN was constructed by common key regulatory elements interacted with common DEG targets. A total of ten genes that were identified as hub genes of LUAD specific PPI network and also present in LUAD specific TRN, were determined as potential biomarkers. Finally, survival analysis was conducted to evaluate their prognostic capabilities and biomarker genes showed significant prognostic performances and high capabilities in classifying risk groups.

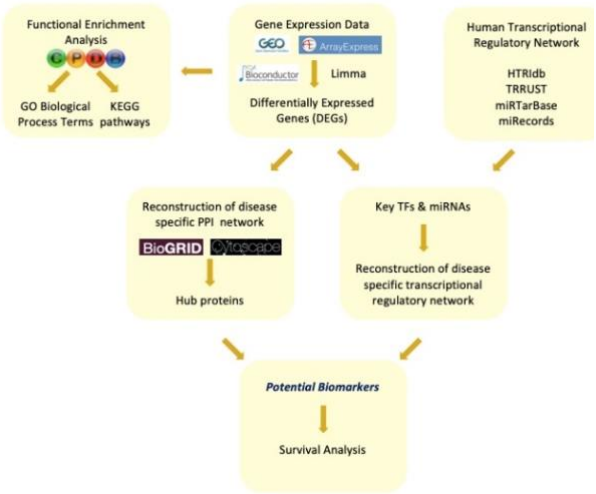

Figure 1. Schematic illustration of the system-based integrative analysis carried out in the present study.

Materials and Methods

Gene expression data collection and analysis

Gene expression datasets from three independent studies [GSE118370 (Xu Liyun et al., 2018), E-MTAB-5231 (Willuda et al., 2017), and GSE40791 (Y. Zhang et al., 2012)] associated with LUAD were selected from ArrayExpress or Gene Expression Omnibus databases. The datasets obtained by Affymetrix Human Genome U133 Plus 2.0 arrays were used to avoid altered gene expressions due to microarray differences. GSE118370 was consisted of tumor and paired adjacent non-tumor tissues from 6 LUAD patients. E-MTAB-5231 contained 22 NSCLC samples, including AD and SCC samples, and 18 normal lung tissues adjacent to tumor samples. Since AD subtype was specifically investigated in this study, SCC samples were excluded from this dataset and data obtained from samples of two stage I and seven stage II LUAD patients were analyzed. GSE40791 contained 94 tumor tissues from 69, 12, and 13 stage I, II, and III LUAD patients, respectively and 100 non-tumor tissues. Two independent datasets, GSE63459 (Robles et al., 2016) containing 33 stage I LUAD tissues and their non-tumor adjacent tissues, and GSE75037 (Girard et al., 2017) containing 83 samples from LUAD tissues and their non-malignant adjacent tissues were used to assess the expression of candidate biomarker genes. By using these two datasets, principal component analysis (PCA) was performed via R software (*R Core Team*, 2020) to elucidate whether the candidate biomarker genes could discriminate LUAD tissues and non-tumoral tissues according to gene expression levels.

Each gene expression dataset was normalized independently and the associated significantly expressed genes were determined using R/Bioconductor (Gentleman et al., 2004). Quantile normalization was performed by RMA (Bolstad et al., 2003) option of the *affy* package (Gautier et al., 2004) and multiple testing option of LIMMA (Smyth, 2004) was used for statistical analysis. The false-discovery rate was controlled by Benjamini-Hochberg's method (Benjamini & Hochberg, 1995). An adjusted *P-value* threshold of 0.05 was used to determine the significance of gene expression and the genes showing at least 2-fold change (FC) in their expression levels were identified. The genes satisfying both *P-value* and FC thresholds, were defined as DEGs. Scatter volcano plots were constructed via R using log2 FCs and corresponding *P-values* of all genes in gene expression data. A heat map that represents gene expression profiles of common DEGs among all samples, was plotted using R. Overrepresentation analyses were performed by ConsensusPathDB (Kamburov et al., 2013), and a *P-value* threshold of 0.05 was used to identify significantly enriched KEGG pathways and GO biological processes associated with DEGs.

Reconstruction of disease-specific protein-protein interaction network

To construct a PPI network of LUAD, physical interactions between the proteins encoded by common DEGs were extracted from BioGrid (Stark et al., 2006) database release 4.2.191. This network was further analyzed and visualized via Cytoscape (Shannon et al., 2003) version 3.7.2. To identify hub proteins, node scores were determined using CytoHubba (Chin et al., 2014). All nodes were ranked based on twelve scoring metrics, including degree, maximal clique centrality (MCC), maximum neighborhood component (MNC), density of maximum neighborhood component (DMNC), betweenness, edge percolated component (EPC), bottleneck, eccentricity, closeness, radiality, clustering coefficient (CC) and stress. Top ten proteins of each scoring metric were isolated and the proteins that were commonly identified by at least five scoring methods were determined as hub proteins.

Reconstruction of disease-specific transcriptional regulatory network

The reported transcription factor (TF)-target gene interactions in HTRIdb (Bovolenta et al., 2012) and TRRUST v2 (Han et al., 2018) databases, as well as microRNA (miRNA)-target gene interactions with strong experimental evidences deposited in miRTarBase release 8.0 (Chou et al., 2018) and miRecords (Xiao et al., 2009) databases were used to reconstruct human TRN. The constructed network was composed of 25669 interactions between 827 TFs and 12659 genes; and 9905 interactions between 844 miRNAs and 3269 genes. For each gene expression dataset, TF-target DEG and miRNA-target DEG interactions were extracted from human TRN. Enrichment analysis were conducted using hypergeometric distribution function, and *P-value* threshold of 0.05 was used to identify key regulatory elements, i.e., TFs and miRNAs, in the presence of LUAD. Disease-specific TRN was constructed by employing key regulatory elements that were found to be common in three datasets interacted with common DEGs.

Identification of potential biomarkers for LUAD

Candidate biomarkers for LUAD were determined by simultaneous analysis of disease-specific networks, i.e., PPI and TRN. Genes that encode hub proteins and are also present in the disease-specific TRN were considered as potential biomarkers for LUAD. The disease involvements of key TFs and biomarker genes as well as their prognostic capabilities in cancer were investigated via Human Protein Atlas available from http://www.proteinatlas.org (Uhlen et al., 2017). Human miRNA and disease associations were identified by the Human microRNA Disease Database (HMDD v 3.2) (Lu et al., 2008).

Survival analysis

The prognostic capabilities of potential biomarkers were assessed using gene expression data containing 475 samples obtained from TCGA. Cox proportional hazards regression analysis was carried out via SurvExpress (Aguirre-Gamboa et al., 2013), and according to the prognostic index, patients were classified as high- and low-risk groups. Gene expression levels of risk groups and survival times were visualized by box-plots and Kaplan-Meier plots, respectively. A log-rank *P-value* cut-off was maintained as 0.05 to define the statistical significance of survival.

Results

Analysis of significantly and differentially expressed genes

The gene expression profiles of tumoral and non-tumoral lung tissues obtained from three independent studies were comparatively analyzed. A total of 1349 genes (425 up, 924 down), 1447 genes (666 up, 781 down), 2202 genes (777 up, 1425 down) were identified as DEGs in GSE118370, E-MTAB-5231, GSE40791, respectively, and scatter volcano plots were plotted to illustrate the distribution of each gene according to the log2FC and –log (*P-value*) values (Figure 2a). Although the numbers of DEGs among datasets were incompatible, 432 DEGs (83 up- and 349 down-regulated) were found to be significantly expressed in common (Figure 2b). A heat map showed the expression profiles of common 432 DEGs identified in the analysis. The common 432 DEGs, including 349 significantly down-regulated genes and 83 significantly up-regulated genes, could effectively distinguish LUAD samples from normal samples (Figure 2c). Since common DEGs might hold important information on LUAD, the significantly enriched GO biological process terms and KEGG pathways (Figure 3a) and protein classes (Figure 3b) associated with these genes were identified.

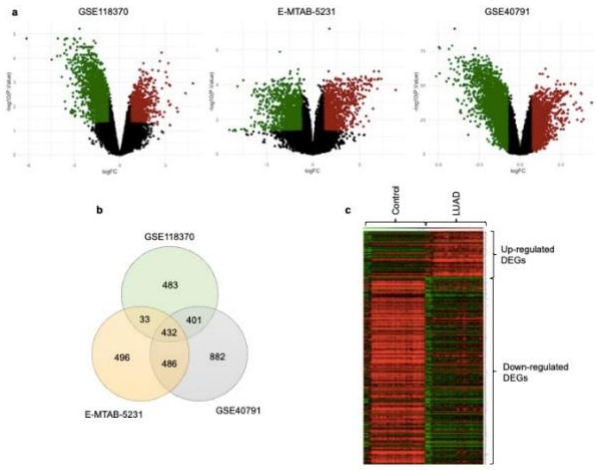

Figure 2. a) The volcano plots for the transcriptome datasets. The x-axis represents the log2 transform of fold change ratios; the y-axis represents the log10 transformed adjusted *P-value*. The red and green colored dots represent the up- and down-regulated DEGs, respectively. **b)** Venn diagram representing the comparison of DEGs in the datasets. **c)** Heatmap representation of common DEGs.

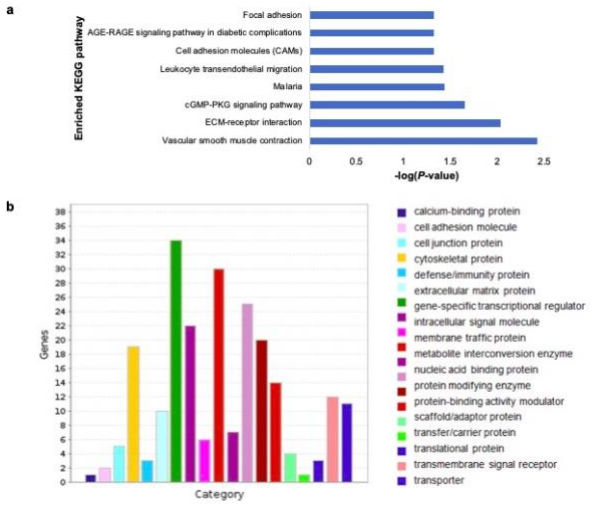

Figure 3. a) Significantly enriched KEGG pathways of common DEGs. **b)** Panther protein classes of common DEGs.

A total of 83 genes, which were commonly induced in LUAD tissues, were found to be significantly associated with cell cycle, DNA damage, cell adhesion, extracellular matrix (ECM) related processes. The top significant processes related to the commonly repressed 349 genes were angiogenesis, cell adhesion, vasculogenesis, and ECM organization. Moreover, pathway analysis indicated vascular smooth muscle contraction, ECM-receptor interaction, cGMP-PKG signaling pathways among the significantly enriched KEGG pathways of 432 common DEGs.

Identification of hub proteins

LUAD specific PPI network constructed by collecting interactions between the proteins encoded by common DEGs, contained 204 interactions between 165 proteins (Figure 4). Hub proteins, which might have important roles in the progression of the disease, were determined by simultaneous investigation of twelve scoring metrics. The proteins that are among the top ten proteins determined by at least five scoring metrics were identified as hub proteins. Aurka, Cav1, Chmp4c, Clu, Eno1, Fhl1, Fhl2, Lmo2, Myh10, Myh11, Nme1 and Sfn were found as hub proteins.

Identification of key regulatory elements

Transcriptional and post-transcriptional regulatory changes in the presence of LUAD were elucidated by integrative analysis of gene expression data and the constructed human TRN. Key regulatory elements for each dataset were determined by extracting TF-DEG and miRNA-DEG interactions from human TRN and calculating hypergeometric probability in the disease state considering a statistical significance threshold of *P-value* < 0.05. Key miRNAs, including miR-103a-3p, miR-135a-5p, miR-200c-3p, miR-203a-3p, miR-204-5p, miR-21, miR-223, miR-25-3p and miR-29a-3p were identified in LUAD (Table 1). Disease associations of key miRNAs were investigated via HMDD and all of them were found to be linked to either lung neoplasms or NSCLC, including LUAD for six of them. Moreover, two of the key miRNAs, miR-200c and miR-21, were previously reported among the significantly expressed miRNAs in stage I LUAD patients and survival analysis indicated that miR-21 was significantly associated with the prognosis of LUAD (Robles et al., 2016). A total of 34 TFs were identified as key regulatory TFs. Top five of the key TFs were also presented in Table 1. The gene products of 25 key TFs were previously determined as prognostic markers of various cancers with Fosl1 being a prognostic marker of lung cancer. Moreover, defects in *IRF1* was associated with lung cancer. TRN of LUAD was constructed by isolating the interactions between key regulatory elements and common DEGs from human TRN. The disease-specific TRN contained 680

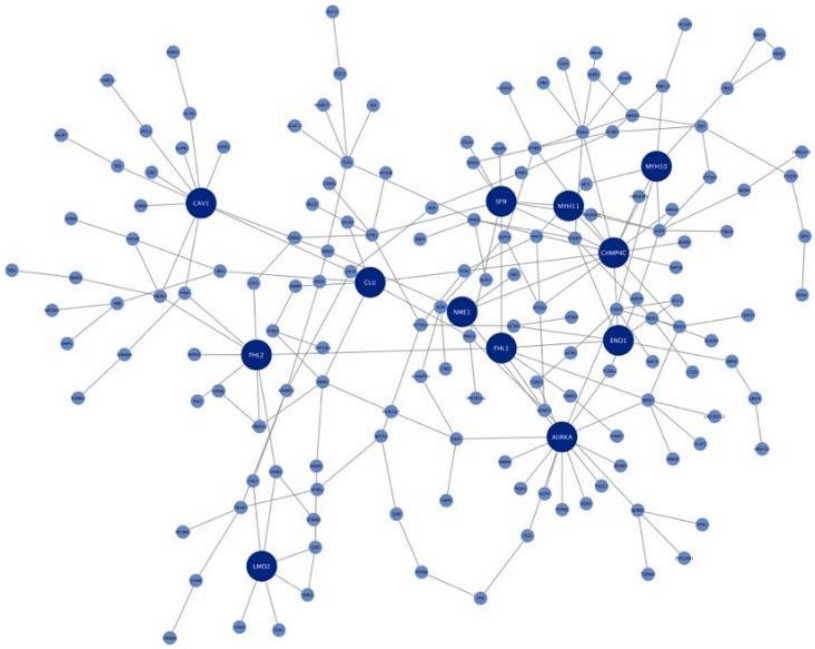

Figure 4. LUAD specific PPI network (dark blue nodes represent hub proteins).

Table 1. Key regulatory elements in LUAD

Key miRNA	Feature
miR-103a-3p	Afflicted with lung neoplasms
miR-135a-5p	Afflicted with lung neoplasms, LUAD
miR-200c-3p	Afflicted with lung neoplasms, lung fibrosis, NSCLC
miR-203a-3p	Afflicted with lung neoplasms, NSCLC, LUAD
miR-204-5p	Afflicted with NSCLC, LUAD
miR-21	Afflicted with lung neoplasms, NSCLC, SCC
miR-223	Afflicted with lung neoplasms, NSCLC, LUAD
miR-25-3p	Afflicted with lung neoplasms, NSCLC, LUAD
miR-29a-3p	Afflicted with lung neoplasms, lung fibrosis, NSCLC, LUAD
Key TFs	**Prognostic marker**
AR (Androgen receptor)	renal cancer (favorable) and liver cancer (favorable)
RELA (RELA proto-oncogene, NF-kB subunit)	renal cancer (unfavorable) and liver cancer (unfavorable)
SP1 (Sp1 transcription factor)	pancreatic cancer (unfavorable)
ESR1 (Estrogen receptor 1)	endometrial cancer (favorable)
YBX1 (Y-box binding protein 1)	renal cancer (unfavorable) and liver cancer (unfavorable)

interactions between 43 key regulatory elements (34 TFs and nine miRNAs) and 306 common DEGs.

Determination of potential biomarkers

Disease-specific PPI network and TRN were used to identify potential biomarkers of LUAD. Genes that encode hub proteins were regarded as potential biomarkers only if they were members of LUAD specific TRN. *AURKA, CAV1, CLU, ENO1, FHL1, FHL2, LMO2, MYH11, NME1* and *SFN* were discovered as biomarkers and their functions, disease involvements and prognostic capabilities in cancer were presented in Table 2.

Biomarker genes were also analyzed for enriched GO biological processes. Cell cycle, DNA damage checkpoint, DNA integrity checkpoint, cell death, apoptotic signaling pathway, cell proliferation, lipid metabolic process, response to lipid, reactive oxygen species biosynthetic process, cardiocyte differentiation, circulatory system development, heart development, and muscle contraction were found among the significantly enriched GO biological processes.

Candidate biomarker genes were also assessed by using two independent datasets, GSE63459 and GSE75037 containing samples from LUAD tissues and their non-tumoral adjacent lung tissues. All biomarker genes were significantly expressed in GSE63459 and the expressions of all biomarker genes, except *ENO1* were found to be significantly altered in GSE75037. The capabilities of biomarker genes to discriminate LUAD and non-tumoral lung tissues were assessed by PCA. PCA was carried out based on the expression profiles of biomarker genes in GSE63459 and GSE7503, and the first three principal components describing at least 74% of the total variance were considered. The biomarker genes as a group showed high potential in discriminating LUAD tissues from non-tumoral adjacent lung tissues (Figure 5).

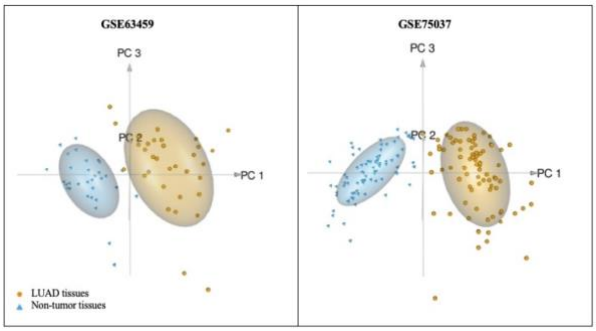

Figure 5. PCA was applied using the gene expression profiles of biomarker genes and the first three principal components could separate LUAD tissues from non-tumoral adjacent lung tissues in GSE63459 and GSE75037. Samples from LUAD tissues, and non-tumoral adjacent lung tissues were presented in yellow and blue colors, respectively. Ellipses indicate 95% confidence interval.

Survival analysis of potential biomarkers

Table 2. Summary of identified biomarker genes

Gene	Prognostic marker
AURKA (Aurora Kinase A)	renal cancer (unfavorable), endometrial cancer (unfavorable), liver cancer (unfavorable) and pancreatic cancer (unfavorable)
CAV1 (Caveolin 1)	lung cancer (unfavorable) and renal cancer (unfavorable)
CLU (Clusterin)	thyroid cancer (favorable)
ENO1 (Enolase 1)	liver cancer (unfavorable) and glioma (unfavorable)
FHL1 (Four and a half LIM domains 1)	urothelial cancer (unfavorable)
FHL2 (Four and a half LIM domains 2)	renal cancer (unfavorable) and head and neck cancer (unfavorable)
LMO2 (LIM domain only 2)	liver cancer (favorable)
MYH11 (Myosin heavy chain 11)	liver cancer (favorable)
NME1 (NME/NM23 nucleoside diphosphate kinase 1)	renal cancer (unfavorable) and liver cancer (unfavorable)
SFN (Stratifin)	renal cancer (unfavorable), liver cancer (unfavorable), endometrial cancer (favorable) and pancreatic cancer (unfavorable)

The prognostic performances of potential biomarkers were determined via Cox proportional hazards regression analysis performed in SurvExpress. Patients were classified as high- and low-risk groups according to prognostic index estimated via expression levels of biomarker genes. Box-plots and Kaplan-Meier plots were used to visualize gene expression levels of risk groups and survival probabilities, respectively (Figure 6). The difference in gene expression between risk groups were compared using t-test and the biomarker genes were found to have significantly different (*P-value* < 0.05) expression levels in high- and low-risk groups (Figure 6a). Moreover, analysis indicated the significant prognostic performance of biomarker genes as a group (Hazard ratio = 1.72 and log-rank *P* = 0.0005) (Figure 6b). Biomarker genes showed high performance in classifying the patients with long and short survival and approximately 1.72-fold difference in death rate between these groups was observed.

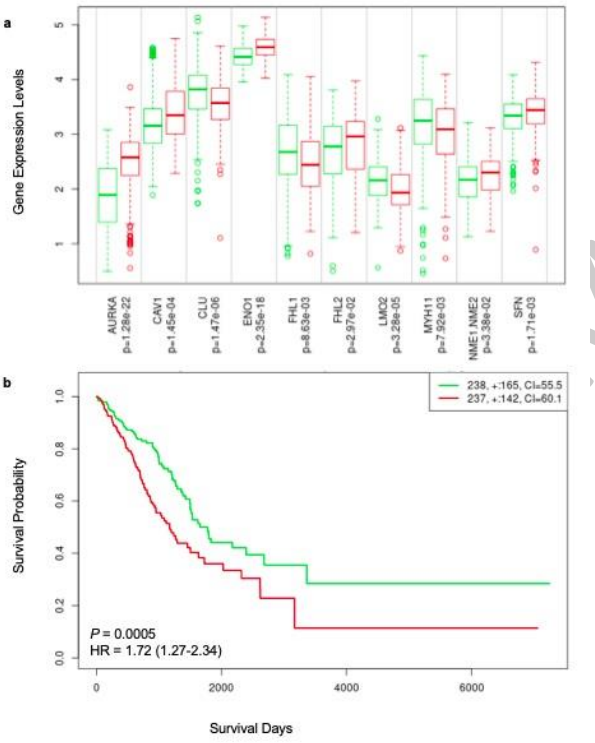

Figure 6. Survival analysis of potential biomarker genes. **a)** Box-plot showing the range and distribution of the gene expression levels of biomarker genes in high- and low-risk groups and *P-value*s representing the differences in biomarker genes' expression levels between high- and low-risk groups. **b)** Kaplan-Meier plot of biomarker genes. Hazard Ratio (HR) was given with a 95% confidence interval. Red and green curves denote high- and low-risk groups, respectively.

Discussion

Lung cancer is the leading cause of cancer associated deaths worldwide (Yan et al., 2019). Due to the high mortality rate, the identification of predictive biomarkers with high prognosis is still an issue. This study developed a computational framework to discover biomarkers for LUAD that is the most common type of lung cancer. The gene expression profiles obtained from healthy and tumoral lung tissues were comparatively analyzed to identify DEGs. Transcriptomics data derived from three independent studies was used to catch diverse processes, pathways, and key molecules that cannot be observed by a single dataset and to reduce the errors associated with a single dataset. Transcriptome, interactome and regulome were used simultaneously for the construction of disease-specific networks, i.e., PPI network and TRN. The integrative analysis of disease-specific networks yielded *AURKA*, *CAV1*, *CLU*, *ENO1*, *FHL1*, *FHL2*, *LMO2*, *MYH11*, *NME1* and *SFN* as potential biomarkers that could be used in the diagnosis and/or prognosis of LUAD and survival analysis revealed their significant prognostic performances.

The analysis of gene expression datasets used in this study revealed 432 common DEGs whose expression patterns were predominantly down-regulated (81%). Functional enrichment analysis of common DEGs revealed three major mechanisms that accompany LUAD: changes in cell adhesion properties, possibility of bone metastasis, and pulmonary vascular remodeling. i) Enrichment analysis indicated cell adhesion, ECM organization and collagen associated processes among the affected processes. Collagen is the most abundant component of ECM, and the modifications in the content and the distribution of collagen due to tumor microenvironment lead to structural changes in ECM during the development of cancer (Lai Xu et al., 2017; S. Xu et al., 2019). Therefore, the induction and repression observed in ECM organization and collagen metabolism were in accordance with the loss of adhesion properties of tumor cells. ii) Bone is one of the most common sites of metastasis and bone metastasis occurs in 30-40% of the advanced lung cancer patients (Macedo et al., 2017). Genes functioning in bone associated processes, such as ossification, fibroblast growth factor, bone remodeling and mineralization were found to be significantly expressed, which might indicate the possible occurrence of bone metastasis in LUAD patients. iii) Significant repression of genes involved in cardiovascular system related processes that might be due to pulmonary vascular remodeling was observed. Pulmonary vascular remodeling frequently accompanies lung cancer and leads to the thickening of blood vessel wall. Since the accumulation of ECM components may cause this type of thickening (Jeffery & Wanstall, 2001; Pullamsetti et al., 2017), the repression of cardiovascular system associated processes together with ECM organization supports the possible changes in the pulmonary vascular structure in LUAD.

Disease-specific networks were constructed by integrating common DEGs with human biological networks. The PPI network was constructed to unveil the interactions among common DEGs and 12 hub proteins were found to be noteworthy in LUAD specific

PPI network. Regulatory elements targeting common DEGs were used for the construction of LUAD specific TRN and key regulatory elements, i.e., 34 TFs and nine miRNAs came into prominence in LUAD. Key miRNAs were previously reported to be associated with lung cancer and key TFs were served as prognostic markers in various cancers (Table 1). Therefore, hub genes regulated by key regulatory elements were assessed as candidate biomarkers for LUAD. Since key regulatory elements might also hold considerable information on the molecular mechanisms of the disease, possible regulatory mechanisms were highlighted and the interactions between key molecules were presented in Figure 7.

When functional enrichment analysis for the discovered biomarkers were conducted, a significant relation between biomarker genes and cell cycle, DNA damage, lipid and cardiovascular system associated processes was observed. The similarity between the processes enriched with biomarkers and that of the common DEGs supported that the discovered biomarkers were representative of LUAD. Therefore, the discovered biomarkers were further investigated to understand their roles in tumor progression. Survival analysis showed that the biomarker genes as a group had high likelihood of being prognostic biomarkers for LUAD. Moreover, the diagnostic capabilities of biomarker genes were evaluated using two independent gene expression datasets and PCA indicated that they could effectively distinguish LUAD samples from normal samples.

Figure 7. Key regulatory elements targeting biomarker genes. Ellipses, rectangles, and triangles represent biomarker genes, TFs, and miRNAs, respectively.

All discovered biomarkers were reported as prognostic markers of various cancers and among them *CAV1* was classified as a prognostic marker of lung cancer (Table 2). *CAV1* encodes a scaffolding protein that is the main component of the caveolae on the plasma membrane. The protein links integrin subunits to the tyrosine kinase FYN, an initiating step in coupling integrins to the Ras-ERK (extracellular signal-regulated kinases) pathway and promoting cell cycle progression. *CAV1* is associated with migration, invasion, and metastasis in cancers and it functions as a tumor suppressor or promoter depending on the stage of the tumor (Shi et al., 2020). Another metastasis suppressor gene, *NME1*, was also detected as a potential biomarker. *NME1* was previously identified because of its reduced mRNA transcript levels in highly metastatic cells. Although its overexpression was not related to the primary tumor size, the metastatic formation was found to be significantly reduced (Marino et al., 2013).

Aurka and Sfn have several functions during mitosis. *AURKA* encodes a cell cycle regulated kinase that is involved in microtubule formation and/or stabilization at the spindle pole during chromosome segregation. It is involved in tumorigenesis through multiple mechanisms and interactions with various proteins functioning as tumor suppressors and oncogenes (Tang et al., 2017), and it has also been linked to poor differentiation of lung cancer (Lo Iacono et al., 2011). *SFN* encodes a cell cycle checkpoint protein that regulates mitotic translation and plays a role in preventing DNA errors during mitosis in response to DNA damage. The suppression of *SFN* expression resulted in a reduction in cell proliferation in LUAD cell line (Shiba-Ishii et al., 2015).

LMO2 encodes a highly conserved cysteine-rich, two LIM-domain protein that has a central and crucial role in hematopoietic development. All human LIM-domain proteins (Lmo1-4) are associated with progression of various cancers. Although *LMO2* is specifically linked to T cell leukaemia, its interacting partner *GAMA2* was involved in the regulation of tumor development in RAS-mutant NSCLC (Matthews et al., 2013). Fhl1 and Fhl2 are also members of the four and a half LIM only protein family. The involvement of *FHL1* and *FHL2* in cancer development is due to their interactions with cancer-related proteins Smad2-4, which results in enhanced expressions of growth inhibiting genes and a decreased expression of a growth promoting gene *c-myc* (Ding et al., 2009). Previous reports also indicated a significant association between the expression of *FHL2* and the cellular level of p53, which is an important tumor suppressor protein (Cao et al., 2015). The communication among the genes encoding these LIM-domain proteins were regulated by *E2F4* that is a common regulator of *LMO2* and *FHL1* together with *AR* that is a common regulator of *FHL1* and *FHL2* in LUAD specific TRN (Figure 7). Since these biomarker genes and their common regulators might play a notable role in LUAD, we suggested them as novel molecular targets in LUAD. Another hub gene, *ENO1* encodes an alpha-enolase and is also involved in growth control and hypoxia tolerance. *ENO1* also modulates *c-myc* expression and inhibits tumor growth (Zhang L. et al., 2018). The expression of *ENO1* was reported to be significantly higher in lung cancer tissues when compared to benign lung disease tissues, however, its expression was not significantly altered to differentiate the subtypes of lung cancer (Zhang L. et al., 2018).

CLU encodes a secreted chaperone that prevents aggregation of non-native proteins and is involved in several basic biological events such as cell death and tumor progression. Intracellular clusterin also regulates the expression of some genes involved in DNA repair. It was previously associated with lung diseases, including asthma and idiopathic pulmonary fibrosis (Habiel et al., 2017). Moreover, its overexpression was observed in a rat model of pulmonary arterial hypertension and contributed to pulmonary vascular remodeling (Liu et al., 2015).

MYH11 encodes a smooth muscle myosin that functions as a major contractile protein, converting chemical energy into mechanical energy through the hydrolysis of ATP. Myosins have several roles in processes related to tumor invasion, such as cell adhesion and migration. *MYH11* was previously reported to be involved in the contraction of airway smooth muscle in asthma. Moreover, the dysregulation of *MYH11* expression was observed in various cancers with a significant decrease in lung cancer (Nie et al., 2020).

Conclusion

There is an ever-growing interest in the identification of diagnostic and prognostic biomarkers and therapeutic targets for lung cancer. Within the framework of this study, the major molecular mechanisms underlying LUAD were determined via comparative transcriptome profiling, and a computational framework was developed to identify potential biomarkers for LUAD. Disease-specific PPI network and TRN were constructed and integrative analysis of these networks with transcriptome data elicited potential biomarkers. The identified biomarkers in this study did not only represent common DEGs functionally, but also showed significant prognostic performance as a group. Possible regulatory communications that might have vital roles in LUAD were also highlighted. Since the identified biomolecules could be valuable for diagnosis and targeted therapies, they deserve clinical investigation. Further experimentation needs to be carried out to verify biomarkers' diagnostic capabilities and to pinpoint their roles in LUAD progression.

References

Aguirre-Gamboa, R., Gomez-Rueda, H., Martínez-Ledesma, E., Martínez-Torteya, A., Chacolla-Huaringa, R., Rodriguez-Barrientos, A., Tamez-Peña, J. G., & Treviño, V. (2013). SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. *PLoS ONE*, *8*(9), 1–9. https://doi.org/10.1371/journal.pone.0074250

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2003). A Comparison of Normalization Methods for High Density Oligonucleotide Array Data based on Variance and Bias. *Bioinformatics*, *19*(2), 185–193. http://www.ncbi.nlm.nih.gov/pubmed/12538238

Bovolenta, L. A., Acencio, M. L., & Lemke, N. (2012). HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*, *13*(405). https://doi.org/10.1186/1471-2164-13-405

Cagle, P. T., Allen, T. C., & Olsen, R. J. (2013). Lung cancer biomarkers: present status and future developments. *Archives of Pathology & Laboratory Medicine*, *137*(9), 1191–1198. https://doi.org/10.5858/arpa.2013-0319-CR

Cao, C. Y., Mok, S. W.-F., Cheng, V. W.-S., & Tsui, S. K.-W. (2015). The FHL2 Regulation in the Transcriptional Circuity of Human Cancers. *Gene*, *572*(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8 (Suppl 4*(511), 1–7. https://doi.org/10.1186/1752-0509-8-S4-S11

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., Huang, W. C., Sun, T. H., Tu, S. J., Lee, W. H., Chiew, M. Y., Tai, C. S., Wei, T. Y., Tsai, T. R., Huang, H. T., Wang, C. Y., Wu, H. Y., Ho, S. Y., Chen, P. R., … & Huang, H. Da. (2018). MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research*, *46*(D1), D296–D302. https://doi.org/10.1093/nar/gkx1067

Cieślik, M., & Chinnaiyan, A. M. (2018). Cancer transcriptome profiling at the juncture of clinical translation. *Nature Reviews Genetics*, *19*(2), 93–109. https://doi.org/10.1038/nrg.2017.96

Ding, L., Wang, Z., Yan, J., Yang, X., Liu, A., Qiu, W., Zhu, J., Han, J., Zhang, H., Lin, J., Cheng, L., Qin, X., Niu, C., Yuan, B., Wang, X., Zhu, C., Zhou, Y., Li, J., Song, H., … & Ye, Q. (2009). Human four-and-a-half LIM family members suppress tumor cell growth through a TGF-β-like signaling pathway. *Journal of Clinical Investigation*, *119*(2), 349–361. https://doi.org/10.1172/JCI35930

Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. a. (2004). Affy-Analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics*, *20*(3), 307–315. https://doi.org/10.1093/bioinformatics/btg405

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., … & Zhang, J. (2004). Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biol*, *5*(10), R80. https://doi.org/10.1186/gb-2004-5-10-r80

Girard, L., Rodriguez-Canales, J., Behrens, C., Thompson, D. M., Botros, I. W., Tang, H., Xie, Y., Rekhtman, N., Travis, W. D., Wistuba, I. I., Minna, J. D., & Gazdar, A. F. (2017). An Expression Signature as an Aid to the Histologic Classification of Non-Small Cell Lung Cancer. *Clin Cancer Res.*, *22*(19), 4880–4889. https://doi.org/10.1158/1078-0432.CCR-15-2900.An

Gov, E., Kori, M., & Arga, K. Y. (2017). Multiomics Analysis of Tumor Microenvironment Reveals Gata2 and miRNA-124-3p as Potential Novel Biomarkers in Ovarian Cancer.

OMICS A Journal of Integrative Biology, *21*(10), 603–615. https://doi.org/10.1089/omi.2017.0115

Habiel, D. M., Camelo, A., Espindola, M., Burwell, T., Hanna, R., Miranda, E., Carruthers, A., Bell, M., Coelho, A. L., Liu, H., Pilataxi, F., Clarke, L., Grant, E., Lewis, A., Moore, B., Knight, D. A., Hogaboam, C. M., & Murray, L. A. (2017). Divergent roles for Clusterin in Lung Injury and Repair. *Scientific Reports*, *7*(15444), 1–14. https://doi.org/10.1038/s41598-017-15670-5

Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C. Y., Lee, M., Kim, E., Lee, S., Kang, B., Jeong, D., Kim, Y., Jeon, H. N., Jung, H., Nam, S., Chung, M., Kim, J. H., & Lee, I. (2018). TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*, *46*(D1), D380–D386. https://doi.org/10.1093/nar/gkx1013

Jeffery, T. K., & Wanstall, J. C. (2001). Pulmonary vascular remodeling: A target for therapeutic intervention in pulmonary hypertension. *Pharmacology and Therapeutics*, *92*(1), 1–20. https://doi.org/10.1016/S0163-7258(01)00157-7

Kamburov, A., Stelzl, U., Lehrach, H., & Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Research*, *41*(Database Issue), D793–D800. https://doi.org/10.1093/nar/gks1055

Li, Z., Sang, M., Tian, Z., Liu, Z., Lv, J., Zhang, F., & Shan, B. (2019). Identification of key biomarkers and potential molecular mechanisms in lung cancer by bioinformatics analysis. *Oncology Letters*, *18*, 4429–4440. https://doi.org/10.3892/ol.2019.10796

Liu, X., Meng, L., Li, J., Meng, J., Teng, X., Gu, H., Hu, S., & Wei, Y. (2015). Secretory clusterin is upregulated in rats with pulmonary arterial hypertension induced by systemic-to-pulmonary shunts and exerts important roles in pulmonary artery smooth muscle cells. *Acta Physiologica*, *213*(2), 505–518. https://doi.org/10.1111/apha.12352

Lo Iacono, M., Monica, V., Saviozzi, S., Ceppi, P., Bracco, E., Papotti, M., & Scagliotti, G. V. (2011). Aurora Kinase A expression is associated with lung cancer histological-subtypes and with tumor de-differentiation. *Journal of Translational Medicine*, *9*(100), 1–6. https://doi.org/10.1186/1479-5876-9-100

Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., & Cui, Q. (2008). An analysis of human microRNA and disease associations. *PLoS ONE*, *3*(10), 1–5. https://doi.org/10.1371/journal.pone.0003420

Macedo, F., Ladeira, K., Pinho, F., Saraiva, N., Bonito, N., Pinto, L., & Gonçalves, F. (2017). Bone metastases: An overview. *Oncology Reviews*, *11*(321), 43–49. https://doi.org/10.4081/oncol.2017.321

Marino, N., Marshall, J.-C., Collins, J. W., Zhou, M., Qian, Y., Veenstra, T., & Steeg, P. S. (2013). Nm23-H1 binds gelsolin and inactivates its actin-severing capacity to promote tumor cell motility and metastasis. *Cancer Res.*, *73*(19), 1–23. https://doi.org/10.1038/jid.2014.371

Matthews, J. M., Lester, K., Joseph, S., & Curtis, D. J. (2013). LIM-domain-only proteins in cancer. *Nature Reviews Cancer*, *13*(2), 111–122. https://doi.org/10.1038/nrc3418

Nie, M. J., Pan, X. T., Tao, H. Y., Xu, M. J., Liu, S. L., Sun, W., Wu, J., & Zou, X. (2020). Clinical and prognostic significance of MYH11 in lung cancer. *Oncology Letters*, *19*(6), 3899–3906. https://doi.org/10.3892/ol.2020.11478

Otálora-Otálora, B. A., Florez, M., López-Kleine, L., Canas Arboleda, A., Grajales Urrego, D. M., & Rojas, A. (2019). Joint Transcriptomic Analysis of Lung Cancer and Other Lung Diseases. *Frontiers in Genetics*, *10*(December), 1–18. https://doi.org/10.3389/fgene.2019.01260

Patel, J. N., Ersek, J. L., & Kim, E. S. (2015). Lung cancer biomarkers, targeted therapies and clinical assays. *Transl Lung Cancer Res*, *4*(5), 503–514.

Pullamsetti, S. S., Kojonazarov, B., Storn, S., Gall, H., Salazar, Y., Wolf, J., Weigert, A., El-Nikhely, N., Ghofrani, H. A., Krombach, G. A., Fink, L., Gattenlöhner, S., Rapp, U. R., Schermuly, R. T., Grimminger, F., Seeger, W., & Savai, R. (2017). Lung cancer-associated pulmonary hypertension: Role of microenvironmental inflammation based on tumor cell-immune cell cross-talk. *Science Translational Medicine*, *9*(416), 1–16. https://doi.org/10.1126/scitranslmed.aai9048

R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/index.html

Robles, A. I., Arai, E., Mathé, E. A., Okayama, H., Schetter, A. J., Brown, D., Petersen, D., Bowman, E. D., Noro, R., Welsh, J. A., Edelman, D. C., Stevenson, H. S., Wang, Y., Tsuchiya, N., Kohno, T., Skaug, V., Mollerup, S., Haugen, A., Meltzer, P. S., … & Harris, C. C. (2016). An Integrated Prognostic Classifier for Stage I Lung Adenocarcinoma based on mRNA, microRNA and DNA Methylation Biomarkers. *J Thorac Oncol*, *10*(7), 1037–1048. https://doi.org/10.1097/JTO.0000000000000560.An

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303.metabolite

Shi, Y. B., Li, J., Lai, X. N., Jiang, R., Zhao, R. C., & Xiong, L. X. (2020). Multifaceted roles of caveolin-1 in lung cancer: A new investigation focused on tumor occurrence, development and therapy. *Cancers*, *12*, 1–19. https://doi.org/10.3390/cancers12020291

Shiba-Ishii, A., Kim, Y., Shiozawa, T., Iyama, S., Satomi, K., Kano, J., Sakashita, S., Morishita, Y., & Noguchi, M. (2015). Stratifin accelerates progression of lung adenocarcinoma at an early stage. *Molecular Cancer*, *14*(142), 1–6. https://doi.org/10.1186/s12943-015-0414-1

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing diferential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*(1), article 3. https://doi.org/10.2202/1544-6115.1027

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*, *34*(Database issue), D535-539. https://doi.org/10.1093/nar/gkj109

Tang, A., Gao, K., Chu, L., Zhang, R., Yang, J., & Zheng, J. (2017). Aurora kinases: Novel therapy targets in cancers. *Oncotarget*, *8*(14), 23937–23954. https://doi.org/10.18632/oncotarget.14893

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., Sanli, K., Von Feilitzen, K., Oksvold, P., Lundberg, E., Hober, S., Nilsson, P., Mattsson, J., Schwenk, J. M., Brunnström, H., … & Ponten, F. (2017). A pathology atlas

of the human cancer transcriptome. *Science*, *357*(eaan2507),1–11. https://doi.org/10.1126/science.aan2507

Villalobos, P., & Wistuba, I. I. (2017). Lung Cancer Biomarkers. *Hematol Oncol Clin North Am.*, *31*(1), 13–29. https://doi.org/10.1016/j.hoc.2016.08.006.Lung

Willuda, J., Linden, L., Lerchen, H. G., Kopitz, C., Stelte-Ludwig, B., Pena, C., Lange, C., Golfier, S., Kneip, C., Carrigan, P. E., McLean, K., Schuhmacher, J., Von Ahsen, O., Müller, J., Dittmer, F., Beier, R., El Sheikh, S., Tebbe, J., Leder, G., … & Kreft, B. (2017). Preclinical antitumor efficacy of BAY 1129980-a novel auristatin-based anti-C4.4A (LYPD3) antibody-drug conjugate for the treatment of non-small cell lung cancer. *Molecular Cancer Therapeutics*, *16*(5), 893–904. https://doi.org/10.1158/1535-7163.MCT-16-0474

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., & Li, T. (2009). miRecords: An integrated resource for microRNA-target interactions. *Nucleic Acids Research*, *37*(Database issue), D105–D110. https://doi.org/10.1093/nar/gkn851

Xu, Lai, Wang, R., Ziegelbauer, J., Wu, W. W., Shen, R. F., Juhl, H., Zhang, Y., Pelosof, L., & Rosenberg, A. S. (2017). Transcriptome analysis of human colorectal cancer biopsies reveals extensive expression correlations among genes related to cell proliferation, lipid metabolism, immune response and collagen catabolism. *Oncotarget*,*8*(43),74703–74719. https://doi.org/10.18632/oncotarget.20345

Xu, Liyun, Lu, C., Huang, Y., Zhou, J., Wang, X., Liu, C., Chen, J., & Le, H. (2018). SPINK1 promotes cell growth and metastasis of lung adenocarcinoma and acts as a novel prognostic biomarker. *BMB Reports*, *51*(12), 648–653. https://doi.org/10.5483/BMBRep.2018.51.12.205

Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T., Zhang, W., Yu, X., & Liu, L. (2019). The role of collagen in cancer: From bench to bedside. *Journal of Translational Medicine*, *17*(1), 1–22. https://doi.org/10.1186/s12967-019-2058-1

Yan, Y., Xu, Z., Qian, L., Zeng, S., Zhou, Y., Chen, X., Wei, J., & Gong, Z. (2019). Identification of CAV1 and DCN as potential predictive biomarkers for lung adenocarcinoma. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, *316*(4), L630–L643. https://doi.org/10.1152/ajplung.00364.2018

Zhang, L., Wang, H., & Dong, X. (2018). Diagnostic value of α-enolase expression and serum α-enolase autoantibody levels in lung cancer. *Jornal Brasileiro de Pneumologia*, *44*(1), 18–23. https://doi.org/10.1590/s1806-37562016000000241

Zhang, Y., Foreman, O., Wigle, D., Kosari, F., Vasmatzis, G., Salisbury, J., van Deursen, J., & Galardy, P. (2012). USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis. *Journal of Clinical Investigation*, *122*(12), 4362–4374. https://doi.org/10.1172/JCI63084DS1