Research article

Identification of potential biomarkers for lung adenocarcinoma

Deepak Sayeeram [a], Teesta V. Katte [a], Saloni Bhatia [a], Anushree Jai Kumar [a], Avinesh Kumar [a], G. Jayashree [a], D.S. Rachana [a], Harsha Vardhan Nalla Reddy [b], Avinash Arvind Rasalkar [c], Rajya Lakshmi Malempati [a], Divijendra Natha Reddy S [a,*]

[a] Department of Biotechnology, BMS College of Engineering, Bengaluru, India
[b] Xe3 Edubioskills Solutions Pvt. Ltd, Bengaluru, India
[c] inDNA Life Sciences Private Limited, Plot 368, 3rd Floor, North View, Infocity Avenue, Patia, Bhubaneswar, Odisha 751024, India

ABSTRACT

Lung adenocarcinoma (LUAD) is the most predominant subtype of lung cancers and is one of the leading causes of cancer related mortality worldwide. Despite the advancements in the field of cancer diagnostics and therapeutics, detection at an early stage using reliable biomarkers is an unmet clinical need for a plethora of cancers, including LUAD, thus attributing to poor prognosis. In view of this, to identify potential biomarkers and therapeutic candidate genes, the expression of all known human genes was screened in the publicly available 'The Cancer Genome Atlas' (TCGA) samples of LUAD patients which resulted in the identification of overexpressed genes. Further analysis of these genes across various patient sample datasets revealed that ZNF687, ODR4, PBXIP1, PYGO2, METTL3, PIGM and RAD1 are consistently more highly expressed in LUAD. Higher expression of these genes either alone or in combination is correlated with poor survival of LUAD patients. Hence, in this study we propose that these identified genes could serve as potential candidates as gene signatures or biomarkers for LUAD that require further investigation in large cohorts of LUAD samples.

1. Introduction

Lung adenocarcinoma (LUAD) is the most common sub-type of non-small cell lung cancer (NSCLC), which is usually detected in the periphery and upper lobes of the lungs, especially in the glandular cells of bronchial mucosa. Amongst the sub-types of NSCLC, LUAD accounts for nearly 40% of all lung cancers and its prevalence is increasing especially among non-smokers (males being at a higher risk than females) and in Asians [1]. Although the initial growth of the cancer occurs at a very early stage, the progression rate is much slower when compared to the other subtypes. However, due to poor prognosis, the overall survival rate is comparatively less than other subtypes [2]. Some studies suggest that a combined exposure to asbestos, cigarette smoke and also certain elements like chromium, nickel and arsenic increases the probability of LUAD development [3, 4]. Secreted Phospho- Protein 1 (SPP1), Ficolin 3 (FCN3), Cyclin-dependent kinase-1 (CDK1) and Mitotic Arrest Deficient 2 Like 1 (MAD2L1) have been shown to be overexpressed in LUAD and are considered as biomarkers due to their consistent expression in the patient samples analyzed from different databases [5]. Recent research suggests that the occurrence of Tumor Infiltrating Leukocytes (TILs)

correlates with the prognosis of lung adenocarcinoma [6]. Given the poor prognosis of LUAD, a large number of studies are warranted towards discovering biomarkers which are highly expressed in patient samples (at various stages from very early to late) across various population demographics and occupations. The aim of this work is to determine biomarkers for LUAD, which may help in early diagnosis and thereby prolonging overall survival of patients. The RNA sequencing data from The Cancer Genome Atlas (TCGA-2014) has been used to identify highly expressed genes in LUAD [7].

2. Materials and methods

2.1. Patient datasets and gene expression analysis

For the analysis of the genes involved in LUAD, the list of 19755 human genes was downloaded from the Hugo Gene Nomenclature Committee (HGNC) (https://www.genenames.org/cgi-bin/download) and its abbreviations were retained for conducting the study [8]. To download, visualize and analyze large datasets, "cBioPortal" (http://www.cbioportal.org/) was used [9, 10]. The TCGA-2014 data set,

consisting of 230 LUAD patient samples including the normal adjacent samples as reference was considered for the analysis [10]. Samples with RNA sequencing data were used to study mRNA expression, mutations and copy number variations in human genes [7]. To identify the differential mRNA expression among the samples, a minimum of two-fold higher was considered as differentially expressed. cBioPortal arrives at mRNA expression data by computing the expression of an individual gene relative to the gene's expression distribution in a reference population. The reference population includes all samples that are diploid for the gene in question (by default for mRNA), or normal samples (when specified), or all profiled samples. The returned value indicates the number of individuals deviating from the mean of expression (standard deviation) in the reference population (Z-score). This measure is useful to determine whether a gene is up- or down-regulated relative to the normal samples or all other tumor samples, and the output is provided as an oncoprint. Oncoprint is a pictorial representation of up or down regulated genes, displaying either red bars (higher expression) or blue (lower expression) in affected individuals. Each bar represents one sample. Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/) an interactive web tool, was used for analyzing RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and the GTEx projects employing standard statistical methods or tools [11]. Four-way analysis of variance (ANOVA), using sex, age, ethnicity and disease state (Tumor or Normal) as variables, is used for calculating the differential expression of the corresponding genes. The expression data are first Log_2 (TPM+1) transformed, and the log_2FC is defined as the difference between median (Tumor) and median (Normal). The Benjamini and Hochberg false discovery rate (FDR) method was used to adjust the p-value across each factor to obtain the multiple testing adjusted q-value. UALCAN (http://ualcan.path.uab.edu/index.html) is a user-friendly online web tool useful for analyzing the TCGA data for gene expression, survival and protein expression [12]. It uses Transcripts Per Million (TPM) as a measure of gene expression and provides box plots by comparing gene expression levels with corresponding normal samples in that data set. It also provides gene expression in different disease stages of LUAD. It utilizes all regular statistical methods of R and uses t-test for estimating p-value.

2.2. Patient survival plots

Kaplan-Meier curves and estimates of survival were determined using SurvExpress (http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp), and the significance was calculated by log-rank test [13]. The correlation was performed between either an individual gene or a group of genes and the survival of the patients. The TCGA-2016 LUAD data consisting of 475 patients was used with death as the stratification parameter. PROGgene V2 (http://genomics.jefferson.edu/proggene/) was used to correlate the over-expression gene cluster with the survival of the patient samples across various data [14].

2.3. Protein expression and gene functional analysis

String database (https://string-db.org) [15] or cBioPortal was used for constructing the network diagram for shortlisted genes. Human Protein Atlas (HPA) (https://www.proteinatlas.org/) was used to obtain antibody-stained images for LUAD tissue samples for various proteins [16]. Protein expression was also obtained for the different stages of LUAD from quantitative proteomics studies of CPTAC samples using the UALCAN web tool (http://ualcan.path.uab.edu/index.html). To determine the gene expression in various stages of the disease, GEPIA or CANCERTOOL (http://web.bioinformatics.cicbiogune.es/CANCERTOOL/) was used [17]. The expression of the genes was studied in various databases [18, 19, 20, 21]. For process and functional analysis, FunRich software (version 3.1.3; 64-bit) was used [22].

3. Result and discussion

3.1. Identification of differentially-expressed genes in LUAD patient samples

In order to determine differentially-expressed genes in LUAD patient samples without any bias, a list of all known human genes (obtained from HGNC) was screened for expression in the patient data sets of LUAD available in TCGA database. To shortlist the genes we chose a minimum of two-fold higher expression in the disease condition. We also
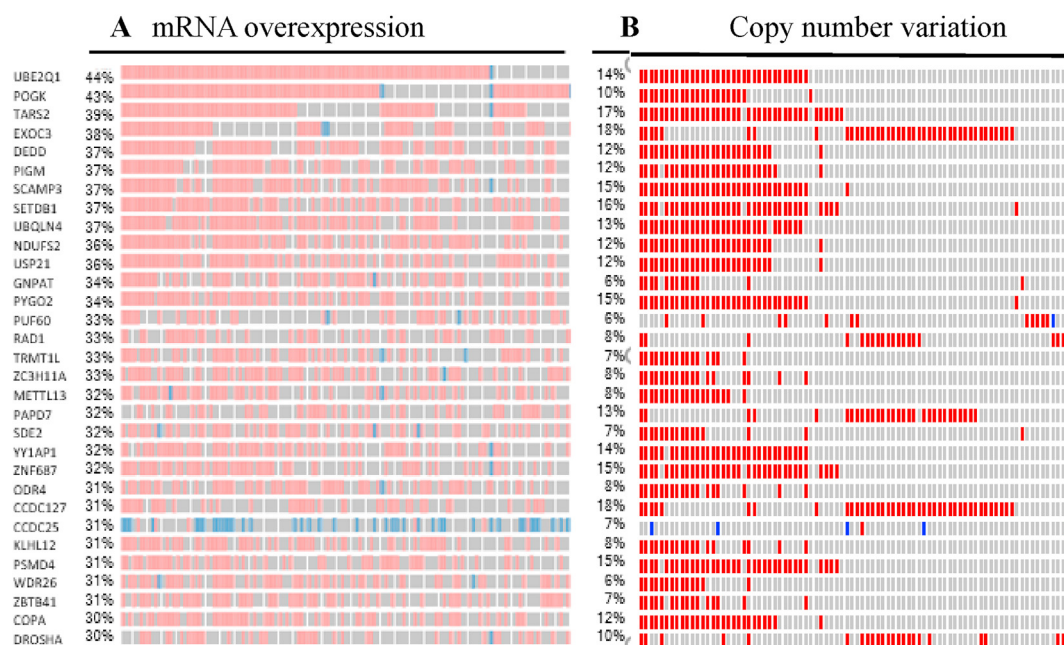

Figure 1. Co-relation between mRNA expression profile and copy number alteration (CNA) of the various genes in LUAD. (A) mRNA upregulation for genes in LUAD patient samples from TCGA 2014 (N = 230). (B) CNA status of same set of genes in the sample dataset. Blue bars indicate lower mRNA levels in A or deletion in B as compared to the normal whereas red bars indicate over expression or amplification. Each bar indicates a patient sample.
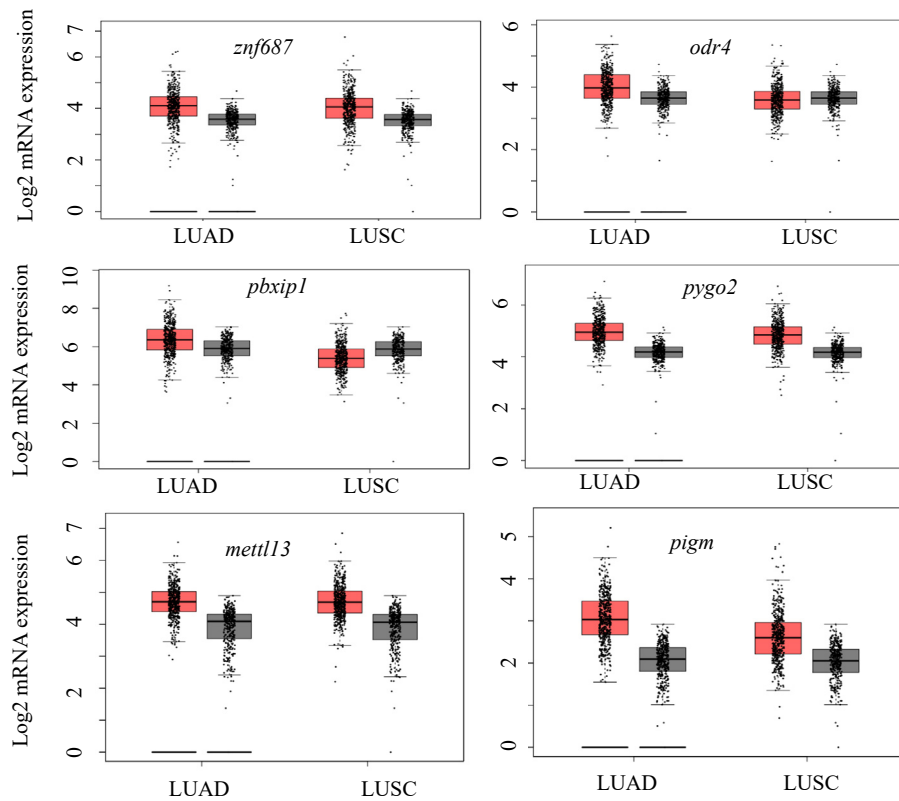
Figure 2. Comparative expression of the genes in LUAD (T = 483, N = 347) and LUSC (T = 486, N = 338). Red and grey boxes indicate the tumor and the normal samples respectively (T = Tumor samples, N=Normal patients). The box plots were obtained from GEPIA.

implemented higher expression in at least 30% of the samples to obtain a manageable number of genes for further analysis. Since multiple studies have suggested that the overexpression of certain genes even in a narrow range patient data set could be considered as a biomarker, we hypothesized that overexpression of the gene in at least 30% of the patient data set could have an immense effect on diagnostic value of LUAD [23, 24, 25]. This resulted in identification of several genes that were expressed two-fold or more in LUAD (Figure 1A). Higher copy number or gene amplification is correlated to the higher expression of genes in many cancers to meet the demands of the tumor physiology and function [26, 27, 28, 29]. Our data also reveals that some of these genes show increased copy number across several patient samples in addition to higher gene expression (Figure 1B). We further validated the higher

expression of these genes in large numbers of LUAD and Lung Squamosal Carcinoma (LUSC) samples. Data show that genes such as *znf687, odr4, pbxip1, pygo2, mettl3,* and *pigm* show the over-expression of mRNA in LUAD samples (Figure 2). Except *odr4* and *pbxip1,* the afore mentioned genes also show high mRNA expression in LUSC (Figure 2). Furthermore, these genes have a higher mRNA expression in LUAD samples in different stages of the disease (Figure 3). As shown in Figure 3, higher expression of these genes is statistically significant in stage 1 LUAD with respect to normal samples, indicating that these could be valuable potential candidates for early stage detection of the disease. These genes also express higher protein levels in LUAD samples in different stages (Figure 4).

We further examined whether these genes are over-expressed in other cancers. Results indicate that *znf687* showed statistically-significant
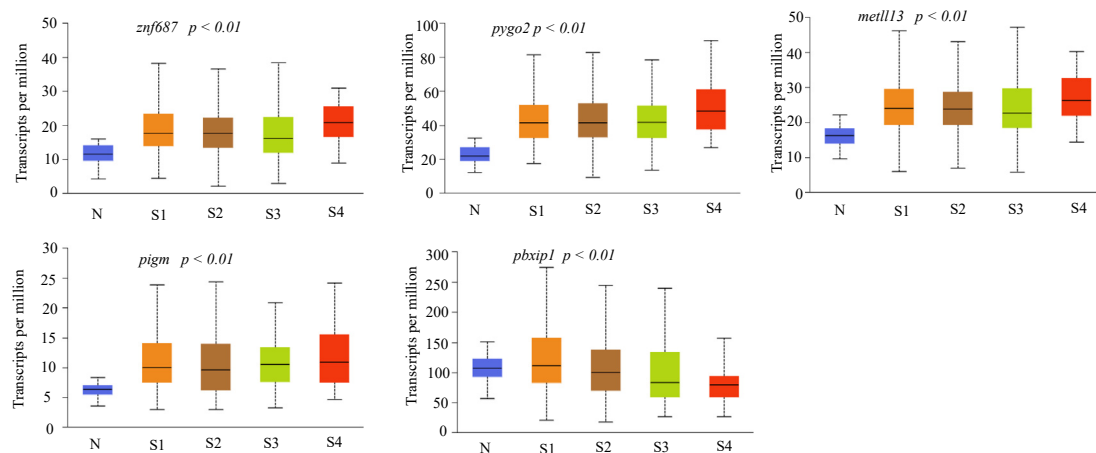

Figure 3. Expression of various genes at different stages of LUAD at mRNA level in TCGA samples. N indicates Normal sample, S1 – S4 indicates different stages of LUAD, having 59, 277, 85 and 28 samples respectively. P < 0.01 is with respect to normal versus stage1 LUAD. The box plots were obtained from UALCAN.
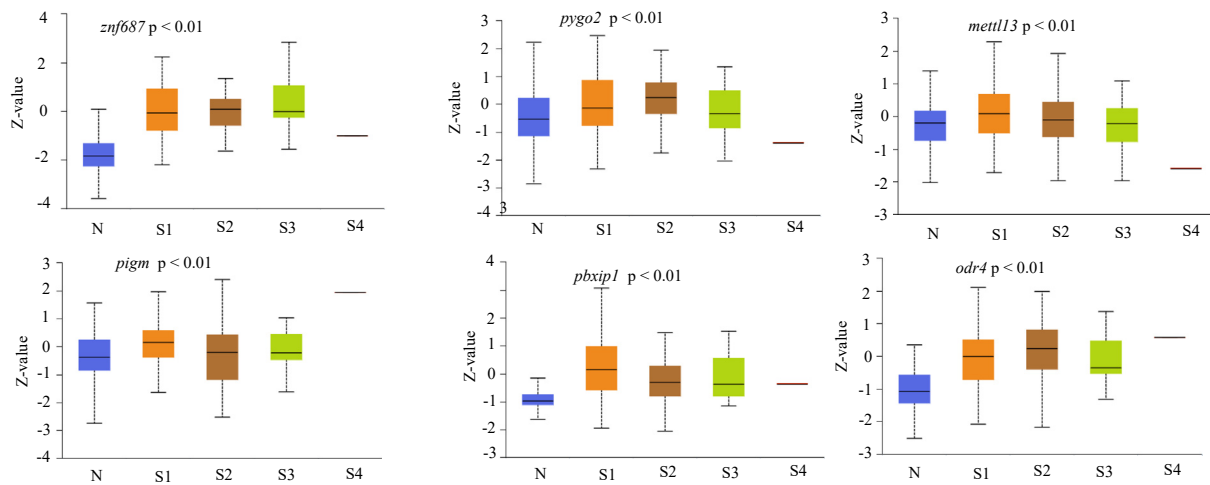
Figure 4. Protein expression level of various genes at different stages of LUAD in CPTAC samples. N indicates Normal sample, S1 – S4 indicates different stages of LUAD having 11, 55, 30, 21, and 1 sample respectively. P-value is with respect to normal versus stage 1 LUAD for all the proteins expression. Z-values represent standard deviations from the median across samples for the given cancer type. The box plots were obtained from UALCAN.

higher expression in esophageal, low grade glioma, liver and thyroid cancers (Figure S1). Similarly, some of the genes, such as *odr4, pbxip1, pygo2, mettl3, and pigm* and *rad1* also show statistically-significant higher expression in many other cancers (Figures S2-S7). To further confirm the expression of these genes at the protein level in LUAD, immunostaining data from Human Protein Atlas (HPA) was analyzed. Some of these genes show higher protein expression status in LUAD in comparison to the normal lung tissue sample (Figure 5) while others show no differential expression between normal and LUAD samples (Figures S8 and 9). Higher protein expression of these genes was further confirmed by

analyzing the CPTAC samples pertaining to LUAD at different stages of the disease (Figure 6). To understand the biological relevance or functions of these genes, FunRich web tool was used. As shown in Figure 7, genes were categorized based on their expression locations, the biological process they operate, and the functional and biological pathways in which they participate. It was observed that most of them are expressed in the cytoplasm or the nucleolus (Figure 7A). Many of these genes are involved in the regulation of cell cycle, metabolism and energy pathways (Figure 7B) which are shown to have significant importance in the context of cancer [30, 31]. When it comes to the functions of these genes,
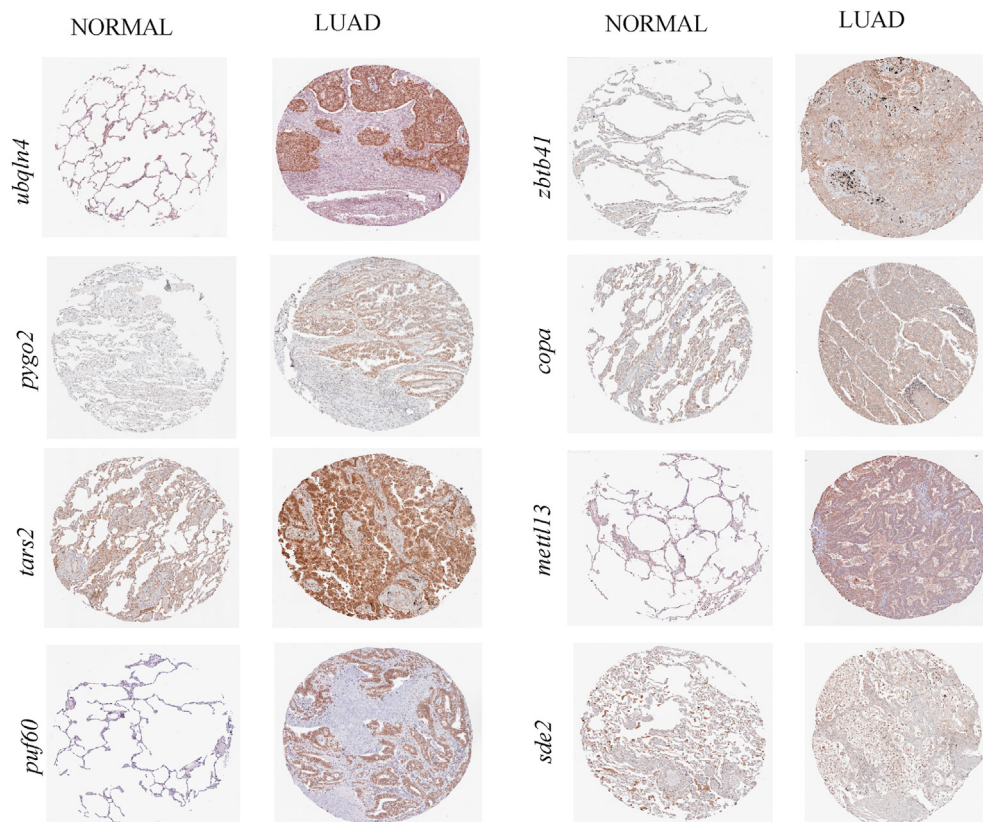

Figure 5. Immunohistochemical staining representing the protein higher expression of the genes in various LUAD tissues. Images were obtained from human protein atlas.
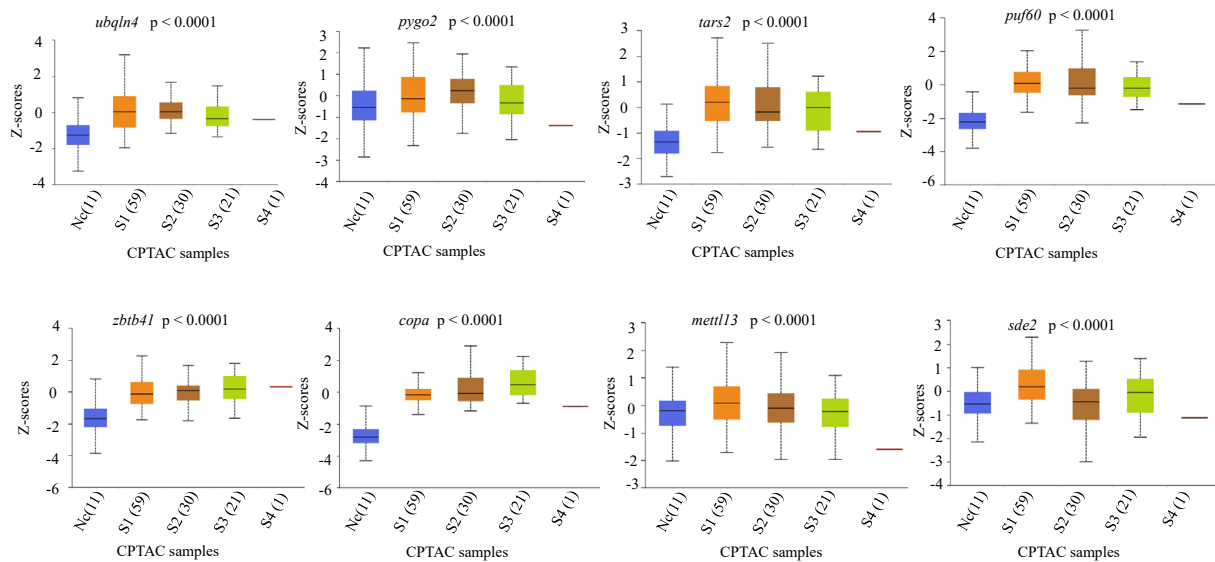
Figure 6. Protein expression in CPTAC (Clinical Proteomic Tumor Analysis Consortium) samples of LUAD. Z-values represent standard deviations from the median across samples for the given cancer type. Log2 Spectral count ratio values from CPTAC were first normalized within each sample profile, then normalized across samples. The stated p values are with respect to normal versus stage 1 and less than 0.05 unless stated otherwise. The box plots were obtained from UALCAN.

many have DNA binding or exonuclease activity (Figure 7C). Many of these genes also operate in the p73 network and cell cycle checkpoints (Figure 7D).

3.2. Patient survival and clinical significance of higher gene expression

To assess the clinical significance of higher expression of these genes on patient survival, SurvExpress containing TCGA data set pertaining to LUAD was used. Results show that higher expression of *znf687, odr4* and *pbxip1* significantly correlate with poor survival (Figure 8A) whereas *mettl3, pigm* and *rad1* did not show a statistically-significant outcome on patient survival. Further analysis showed that higher expression of some of these genes in combination has a statistically-significant effect on patient survival outcome (Figure 8B, C and D). It is interesting to note that genes such as *dedd, scamp3, setdb1, ubqln4, ndufs2, gnpat, rad1, zc3h11a,* and *klhl12* are common in data sets where patient survival is significantly compromised by their higher expression as shown in

Figures 8B, C and D. We also analyzed whether any of these genes show a stage-wise expression pattern in LUAD and thereby serve the utility of being stage- specific diagnostic markers. Violin plots for *pbxip1* and *mettl3* from various LUAD data sets show that these two genes are expressed more during stages 1 and 1A of LUAD (Figure 9A and B). Higher expression of *znf687* is correlated with poor survival of patients in stage 3B of LUAD (Figure S10).

Since, *pbxip1* and *znf687* showed common results across several databases, we wanted to identify their interacting partners and verify whether they could be used as signatures for LUAD (Figure S11). Also, gene expression profile from TCGA LUAD 2014 data revealed that some of the genes that are networked to *pbxip1* and *znf687* show higher mRNA expression in the LUAD (Figure S12). Furthermore, higher expression of *pbxip1* and *znf687* network genes also have statistically-significant poor survival among the affected patients (Figures S13-S14). It has been reported that overexpression of *znf687* enhances tumorigenicity in hepatocellular carcinoma [32].
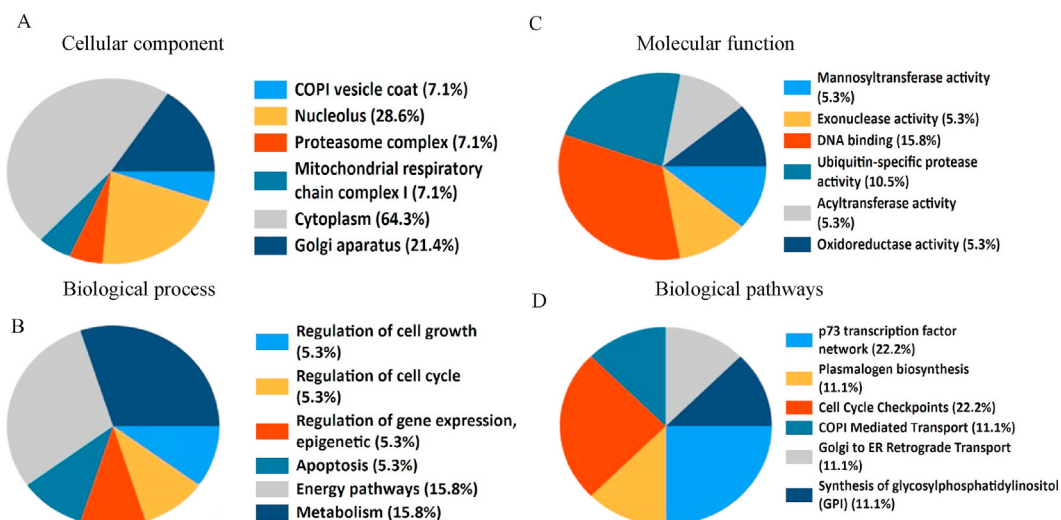

Figure 7. Gene set enrichment analysis. A, B, C, D LUAD higher expressed genes were analyzed with Funrich software for their cellular component, biological process, molecular function and biological pathways.
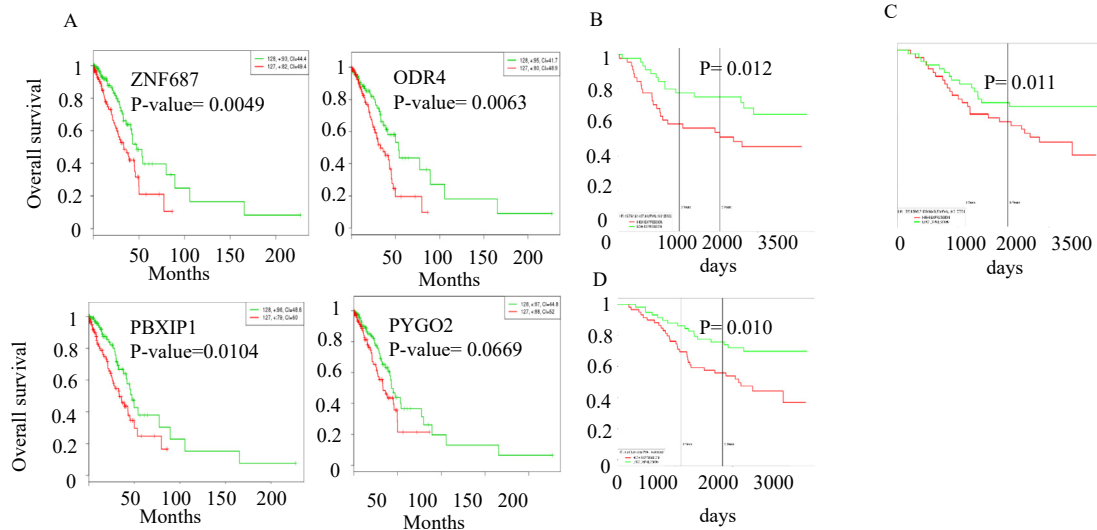
Figure 8. Clinical significance of the higher expressed gene on the patient survival in LUAD (N = 475). (A) Genes that are significantly expressed in LUAD. (B) Prognostic and Predictive Gene Signature for Adjuvant Chemotherapy in Resected Non-Small Cell Lung Cancer for genes *dedd, scamp3, setdb1, ubqln4, ndufs2, usp21, gnpat, puf60, rad1, zc3h11a, mettl13, papd7, yy1ap1, ccdc25, klhl1* (C) Effect on the survival of the diseased individuals by the overexpressed genes in Expression Profile-Defined Classification of Lung Adenocarcinoma data set. for genes *pogk, dedd, pigm, scamp3, setdb1, ndufs2, usp21, gnpat, pygo2, rad1, klhl12, psmd4, wdr26, copa* (D) Effect on the survival of individuals by the overexpressed genes such as *scamp3 setdb1, ubqln, ndufs2, usp21, gnpat, pygo2, rad1, zc3h11a, yy1ap1, znf687, ccdc127, ccdc25, klhl12* in Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis dataset. Green line and red line indicates survival status of the patient samples in which genes are under expressed and over expressed respectively compared to the normal sample. The survival plots were obtained from PROGgeneV2.

4. Discussion

In our study, we identified several over-expressed genes in LUAD and validated them with various bioinformatic tools in LUAD patient samples. It would be interesting to know how many of these genes are already known or identified as potential candidate biomarkers. From the short-listed genes as indicated in Figure 1, genes such as *odr4, pbxip1, mettl3, pigm* and *rad1* are identified as being upregulated in LUAD in this study

and have not been reported as markers for LUAD. Interestingly, all these genes show higher expression at the level of mRNA and protein. Genes such as *setdb1* and *ndufs2* are shown to be upregulated in lung cancer [33, 34], whereas other genes are known to be up/downregulated or are employed as biomarkers for many other cancer types as indicated in the supplementary tables. Nevertheless, all of these genes show higher expression in LUAD patient samples either at the level of mRNA/or protein or both. Early detection and diagnosis of cancers is of tremendous
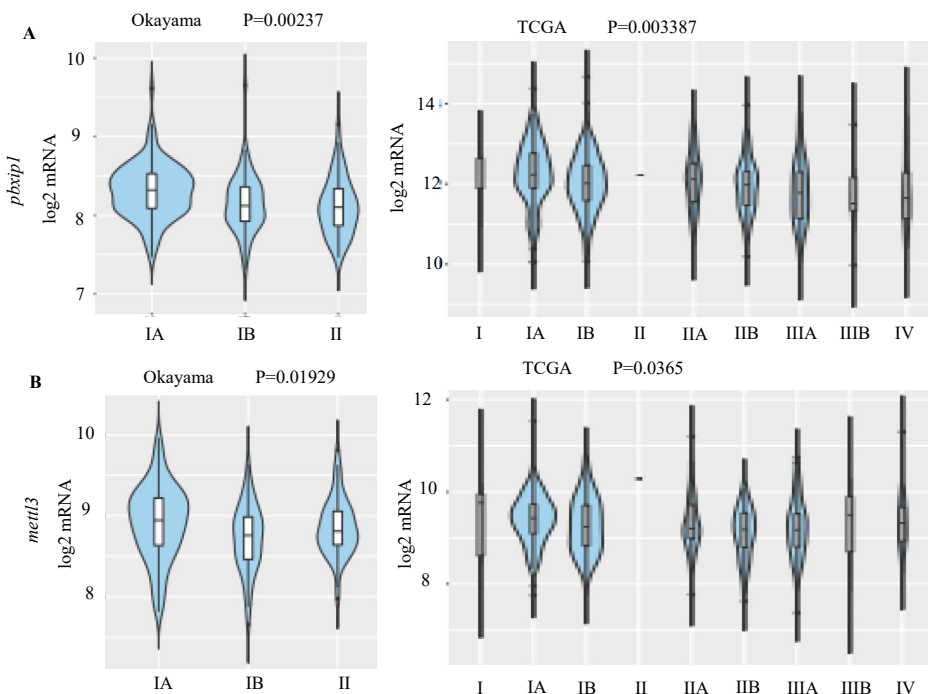

Figure 9. mRNA expression level of a) *pbxip1* and b) *mettl13* in the different stages of LUAD in different datasets - OKAYAMA (N = 246) and TCGA (N = 514). A p-value of 0.05 or less is considered as significant. Violin plots were obtained from CANCERTOOL.

importance in making appropriate decisions for treatment and for better prognosis. Unfortunately, it is still an unmet clinical need that has to be addressed. In this study, genes have been identified which have a statistically significant higher expression pattern in LUAD (stagewise) with respect to normal samples.

There are several approaches/mathematical models that are being employed for determining biomarkers for several diseases. One among them is the Dynamical Network Biomarkers (DNB) and as per this approach, each individual may not have exactly the same DNB even for the same disease, i.e. the markers could vary from person to person [35]. Recent studies have reported potential biomarkers using, single cell RNA sequencing studies in samples either derived from xenografts or patient samples for many cancers and reported potential biomarkers [36]. However, there is no correlation or common overlapping genes as biomarkers between our study and DNB studies pertaining to lung cancer due to the difference in approaches. Identification of reliable biomarkers with high sensitivity and specificity is highly desirable for better management of many cancers [37]. Although there are several research reports on biomarkers for LUAD available, these reports have limited application potential. It would be interesting to test if any of the genes that show higher expression in our study are potential candidates to be used as reliable biomarkers either alone or in combinations and as gene signatures for LUAD. Since a significant proportion of the patient's samples shows higher expression of these genes, we believe that further validation of these genes in a large cohort of patient samples is warranted for developing a suitable and dependable biomarker for LUAD.

Declarations

Author contribution statement

Deepak Sayeeram: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Teesta V Katte: Performed the experiments; Wrote the paper.

Saloni Bhatia, Anushree Jai Kumar, Avinesh Kumar, Jayashree G, Rachana D.S, Avinash Arvind Rasalkar: Performed the experiments.

Nalla Reddy Harsha Vardhan: Performed the experiments; Analyzed and interpreted the data.

Rajya Lakshmi Malempati: Analyzed and interpreted the data; Wrote the paper.

Divijendra Natha Reddy S: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

Divijendra Natha Reddy S was supported by a grant ECR/2016-001685 from Department of Science and Technology, Govt. of India.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2020.e05452.

References

[1] W.D. Travis, E. Brambilla, M. Noguchi, A.G. Nicholson, K.R. Geisinger, Y. Yatabe, D.G. Beer, C.A. Powell, G.J. Riely, P.E. Van Schil, K. Garg, J.H. Austin, H. Asamura, V.W. Rusch, F.R. Hirsch, G. Scagliotti, T. Mitsudomi, R.M. Huber, Y. Ishikawa, J. Jett, M. Sanchez-Cespedes, J.P. Sculier, T. Takahashi, M. Tsuboi, J. Vansteenkiste, I. Wistuba, P.C. Yang, D. Aberle, C. Brambilla, D. Flieder, W. Franklin, A. Gazdar, M. Gould, P. Hasleton, D. Henderson, B. Johnson, D. Johnson, K. Kerr, K. Kuriyama, J.S. Lee, V.A. Miller, I. Petersen, V. Roggli, R. Rosell, N. Saijo, E. Thunnissen, M. Tsao, D. Yankelewitz, International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma, J. Thorac. Oncol. 6 (2011) 244–285.

[2] Z. Hu, X. Gu, R. Zhong, H. Zhong, Tumor-infiltrating CD45RO+ memory cells correlate with favorable prognosis in patients with lung adenocarcinoma, J. Thorac. Dis. 10 (2018) 2089–2099.

[3] F. Mollo, E. Pira, G. Piolatto, D. Bellis, P. Burlo, A. Andreozzi, S. Bontempi, E. Negri, Lung adenocarcinoma and indicators of asbestos exposure, Int. J. Canc. 60 (1995) 289–293.

[4] C.Y. Kuo, R.H. Wong, J.Y. Lin, J.C. Lai, H. Lee, Accumulation of chromium and nickel metals in lung tumors from lung cancer patients in Taiwan, J. Toxicol. Environ. Health Part A Curr. Issues 69 (2006) 1337–1344.

[5] Y.-X. Shi, T. Zhu, T. Zou, W. Zhuo, Y.-X. Chen, M.-S. Huang, W. Zheng, C.-J. Wang, X. Li, X.-Y. Mao, W. Zhang, H.-H. Zhou, J.-Y. Yin, Z.-Q. Liu, Prognostic and predictive values of CDK1 and MAD2L1 in lung adenocarcinoma, Oncotarget 7 (2016) 85235–85243.

[6] X. Hu, Y.Q. Li, Q.G. Li, Y.L. Ma, J.J. Peng, S.J. Cai, ITGAE defines CD8+ tumor-infiltrating lymphocytes predicting a better prognostic survival in colorectal cancer, EBioMedicine 35 (2018) 178–188.

[7] E.A. Collisson, J.D. Campbell, A.N. Brooks, A.H. Berger, W. Lee, J. Chmielecki, D.G. Beer, L. Cope, C.J. Creighton, L. Danilova, L. Ding, G. Getz, P.S. Hammerman, D. Neil Hayes, B. Hernandez, J.G. Herman, J.V. Heymach, I. Jurisica, R. Kucherlapati, D. Kwiatkowski, M. Ladanyi, G. Robertson, N. Schultz, R. Shen, R. Sinha, C. Sougnez, M.-S. Tsao, W.D. Travis, J.N. Weinstein, D.A. Wigle, M.D. Wilkerson, A. Chu, A.D. Cherniack, A. Hadjipanayis, M. Rosenberg, D.J. Weisenberger, P.W. Laird, A. Radenbaugh, S. Ma, J.M. Stuart, L. Averett Byers, S.B. Baylin, R. Govindan, M. Meyerson, M. Rosenberg, S.B. Gabriel, K. Cibulskis, C. Sougnez, J. Kim, C. Stewart, L. Lichtenstein, E.S. Lander, M.S. Lawrence, G. Getz, C. Kandoth, R. Fulton, L.L. Fulton, M.D. McLellan, R.K. Wilson, K. Ye, C.C. Fronick, C.A. Maher, C.A. Miller, M.C. Wendl, C. Cabanski, L. Ding, E. Mardis, R. Govindan, C.J. Creighton, D. Wheeler, M. Balasundaram, Y.S.N. Butterfield, R. Carlsen, A. Chu, E. Chuah, N. Dhalla, R. Guin, C. Hirst, D. Lee, H.I. Li, M. Mayo, R.A. Moore, A.J. Mungall, J.E. Schein, P. Sipahimalani, A. Tam, R. Varhol, A. Gordon Robertson, N. Wye, N. Thiessen, R.A. Holt, S.J.M. Jones, M.A. Marra, J.D. Campbell, A.N. Brooks, J. Chmielecki, M. Imielinski, R.C. Onofrio, E. Hodis, T. Zack, C. Sougnez, E. Helman, C. Sekhar Pedamallu, J. Mesirov, A.D. Cherniack, G. Saksena, S.E. Schumacher, S.L. Carter, B. Hernandez, L. Garraway, R. Beroukhim, S.B. Gabriel, G. Getz, M. Meyerson, A. Hadjipanayis, S. Lee, H.S. Mahadeshwar, A. Pantazi, A. Protopopov, X. Ren, S. Seth, X. Song, J. Tang, L. Yang, J. Zhang, P.-C. Chen, M. Parfenov, A. Wei Xu, N. Santoso, L. Chin, P.J. Park, R. Kucherlapati, K.A. Hoadley, J. Todd Auman, S. Meng, Y. Shi, E. Buda, S. Waring, U. Veluvolu, D. Tan, P.A. Mieczkowski, C.D. Jones, J.V. Simons, M.G. Soloway, T. Bodenheimer, S.R. Jefferys, J. Roach, A.P. Hoyle, J. Wu, S. Balu, D. Singh, J.F. Prins, J.S. Marron, J.S. Parker, D. Neil Hayes, C.M. Perou, J. Liu, L. Cope, L. Danilova, D.J. Weisenberger, D.T. Maglinte, P.H. Lai, M.S. Bootwalla, D.J. Van Den Berg, T. Triche Jr., S.B. Baylin, P.W. Laird, M. Rosenberg, L. Chin, J. Zhang, J. Cho, D. DiCara, D. Heiman, P. Lin, W. Mallard, D. Voet, H. Zhang, L. Zou, M.S. Noble, M.S. Lawrence, G. Saksena, N. Gehlenborg, H. Thorvaldsdottir, J. Mesirov, M.-D. Nazaire, J. Robinson, G. Getz, W. Lee, B. Arman Aksoy, G. Ciriello, B.S. Taylor, G. Dresdner, J. Gao, B. Gross, V.E. Seshan, M. Ladanyi, B. Reva, R. Sinha, S. Onur Sumer, N. Weinhold, N. Schultz, C. Sander, S. Ng, S. Ma, J. Zhu, A. Radenbaugh, J.M. Stuart, C.C. Benz, C. Yau, D. Haussler, P.T. Spellman, M.D. Wilkerson, J.S. Parker, K.A. Hoadley, P.K. Kimes, D. Neil Hayes, C.M. Perou, B.M. Broom, J. Wang, Y. Lu, P. Kwok Shing Ng, L. Diao, L. Averett Byers, W. Liu, J.V. Heymach, C.I. Amos, J.N. Weinstein, R. Akbani, G.B. Mills, E. Curley, J. Paulauskis, K. Lau, S. Morris, T. Shelton, D. Mallery, J. Gardner, R. Penny, C. Saller, K. Tarvin, W.G. Richards, R. Cerfolio, A. Bryant, D.P. Raymond, N.A. Pennell, C. Farver, C. Czerwinski, L. Huelsenbeck-Dill, M. Iacocca, N. Petrelli, B. Rabeno, J. Brown, T. Bauer, O. Dolzhanskiy, O. Potapova, D. Rotin, O. Voronina, E. Nemirovich-Danchenko, K.V. Fedosenko, A. Gal, M. Behera, S.S. Ramalingam, G. Sica, D. Flieder, J. Boyd, J. Weaver, B. Kohl, D. Huy Quoc Thinh, G. Sandusky, H. Juhl, E. Duhig, P. Illei, E. Gabrielson, J. Shin, B. Lee, K. Rogers, D. Trusty, M.V. Brock, C. Williamson, E. Burks, K. Rieger-Christ, A. Holway, T. Sullivan, D.A. Wigle, M.K. Asiedu, F. Kosari, W.D. Travis, N. Rekhtman, M. Zakowski, V.W. Rusch, P. Zippile, J. Suh, H. Pass, C. Goparaju, Y. Owusu-Sarpong, J.M.S. Bartlett, S. Kodeeswaran, J. Parfitt, H. Sekhon, M. Albert, J. Eckman, J.B. Myers, R. Cheney, C. Morrison, C. Gaudioso, J.A. Borgia, P. Bonomi, M. Pool, M.J. Liptay, F. Moiseenko, I. Zaytseva, H. Dienemann, M. Meister, P.A. Schnabel, T.R. Muley, C. Gomez-Fernandez, L. Herbert, S. Egea, M. Huang, L.B. Thorne, L. Boice, A. Hill Salazar, W.K. Funkhouser, W. Kimryn Rathmell, R. Dhir, S.A. Yousem, S. Dacic, F. Schneider, J.M. Siegfried, R. Hajek, M.A. Watson, S. McDonald, B. Meyers, B. Clarke, I.A. Yang, K.M. Fong, L. Hunter, M. Windsor, R.V. Bowman, S. Peters, I. Letovanec, K.Z. Khan, M.A. Jensen, E.E. Snyder, D. Srinivasan, A.B. Kahn, J. Baboud, D.A. Pot, K.R. Mills Shaw, M. Sheth, T. Davidsen, J.A. Demchok, L. Yang, Z. Wang, R. Tarnuzzer, J. Claude Zenklusen, B.A. Ozenberger, H.J. Sofia, W.D. Travis, R. Cheney, B. Clarke, S. Dacic, E. Duhig, W.K. Funkhouser, P. Illei, C. Farver, N. Rekhtman, G. Sica, J. Suh, M.-S. Tsao, Comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (2014) 543–550.

[8] E. Bruford, M.J. Lush, M.W. Wright, T.P. Sneddon, S. Povey, E. Birney, The HGNC database in 2008: a resource for the human genome, Nucleic Acids Res. 36 (2008) 445–448.

[9] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Canc. Discov. 2 (2012) 401–404.

[10] E.L. Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, N.S. Ethan Cerami, Chris Sander, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (2013) 1–34.

[11] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, Z. Zhang, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, Nucleic Acids Res. 45 (2017) W98–W102.

[12] D.S. Chandrashekar, B. Bashel, S.A.H. Balasubramanya, C.J. Creighton, I. Ponce-Rodriguez, B.V.S.K. Chakravarthi, S. Varambally, UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, Neoplasia (United States) 19 (2017) 649–658.

[13] R. Aguirre-Gamboa, H. Gomez-Rueda, E. Martínez-Ledesma, A. Martínez-Torteya, R. Chacolla-Huaringa, A. Rodriguez-Barrientos, J.G. Tamez-Peña, V. Treviño, SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis, PloS One 8 (2013) 1–9.

[14] C.P. Goswami, H. Nakshatri, PROGgeneV2: Enhancements on the Existing Database, 2014, pp. 1–6.

[15] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, J. Kuhn, P. Bork, L.J. Jensen, C. Von Mering, STRING v10: protein – protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (2015) 447–452.

[16] M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Ponten, Tissue-based map of the human proteome, Science (80-.) 347 (2015) 1260419.

[17] A.R. Cortazar, V. Torrano, N. Martín-martín, A. Caro-maldonado, L. Camacho, I. Hermanova, E. Guruceaga, L.F. Lorenzo-martín, V. Quesada, CANCERTOOL: a visualization and representation interface to exploit cancer datasets, Canc. Res. 78 (2018) 6320–6329.

[18] M.D. Wilkerson, Y. Xin, V. Walter, N. Zhao, C.R. Cabanski, M.C. Hayward, C.R. Miller, M.A. Socinski, A.M. Parsons, L.B. Thorne, B.E. Haithcock, N.K. Veeramachaneni, W.K. Funkhouser, S.H. Randell, P.S. Bernard, C.M. Perou, D.N. Hayes, Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation, PloS One 7 (2012).

[19] D. Chitale, Y. Gong, B.S. Taylor, S. Broderick, C. Brennan, R. Somwar, B. Golas, L. Wang, N. Motoi, J. Szoke, J.M. Reinersman, J. Major, C. Sander, V.E. Seshan, M.F. Zakowski, V. Rusch, W. Pao, W. Gerald, M. Ladanyi, An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR -mutant tumors, Oncogene 28 (2009) 2773–2783.

[20] H. Okayama, T. Kohno, Y. Ishii, Y. Shimada, K. Shiraishi, R. Iwakawa, Identification of genes upregulated in ALK -positive and EGFR/KRAS/ALK -negative lung, Adenocarcinomas 72 (2012) 100–112.

[21] K. Shedden, J.M.G. Taylor, S.A. Enkemann, M. Tsao, T.J. Yeatman, W.L. Gerald, S. Eschrich, I. Jurisica, T.J. Giordano, D.E. Misek, A.C. Chang, C.Q. Zhu, D. Strumpf, S. Hanash, F.A. Shepherd, K. Ding, L. Seymour, K. Naoki, N. Pennell, B. Weir, R. Verhaak, C. Ladd-acosta, T. Golub, M. Gruidl, A. Sharma, J. Szoke, M. Zakowski, V. Rusch, M. Kris, A. Viale, N. Motoi, W. Travis, B. Conley, V.E. Seshan,

M. Meyerson, R. Kuick, K.K. Dobbin, T. Lively, Gene expression – based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study,, Nat. Med. 14 (2008) 822–827.

[22] M. Pathan, S. Keerthikumar, D. Chisanga, R. Alessandro, C. Ang, SHORT COMMUNICATION A novel community driven software for functional enrichment analysis of extracellular vesicles data, J. Extracell. Vesicles 6 (2017).

[23] C.Y. Weng, X.Y. Hu, Y.J. Wang, Integrated analysis of gene expression, alteration and clinical significance of carcinoembryonic antigen-related cell adhesion molecule 1 in cancer, 3 Biotech 10 (2020) 1–19.

[24] S. Lin, C. Cao, Y. Meng, P. Wu, P. Gao, W. Zhi, T. Peng, P. Wu, L. Gui, Comprehensive analysis of the value of RAB family genes in prognosis of breast invasive carcinoma, Biosci. Rep. 40 (2020) 1–14.

[25] S.K. Saha, S.M.R. Islam, K.S. Kwak, M.S. Rahman, S.G. Cho, PROM1 and PROM2 expression differentially modulates clinical prognosis of cancer: a multiomics analysis, Canc. Gene Ther. 27 (2020) 147–167.

[26] L. Yang, Y.Z. Wang, H.H. Zhu, Y. Chang, L. Di Li, W.M. Chen, L.Y. Long, Y.H. Zhang, Y.R. Liu, J. Lu, Y.Z. Qin, PRAME gene copy number variation is related to its expression in multiple myeloma, DNA Cell Biol. 36 (2017) 1099–1107.

[27] Y.S. Huang, W. Bin Liu, F. Han, J.T. Yang, X.L. Hao, H.Q. Chen, X. Jiang, L. Yin, L. Ao, Z.H. Cui, J. Cao, J.Y. Liu, Copy number variations and expression of MPDZ are prognostic biomarkers for clear cell renal cell carcinoma, Oncotarget 8 (2017) 78713–78725.

[28] A. Kuzyk, S. Booth, C. Righolt, S. Mathur, J. Gartner, S. Mai, MYCN overexpression is associated with unbalanced copy number gain, altered nuclear location, and overexpression of chromosome arm 17q genes in neuroblastoma tumors and cell lines, Gene Chromosome Canc. 54 (2015).

[29] X. Shao, N. Lv, J. Liao, J. Long, R. Xue, N. Ai, D. Xu, X. Fan, Copy number variation is highly correlated with differential gene expression: a pan-cancer study, BMC Med. Genet. 20 (2019) 1–14.

[30] Z. Xiao, Z. Dai, J.W. Locasale, Metabolic landscape of the tumor microenvironment at single cell resolution, Nat. Commun. 10 (2019) 1–12.

[31] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674.

[32] T. Zhang, Y. Huang, W. Liu, W. Meng, H. Zhao, Q. Yang, S.-J. Gu, C.-C. Xiao, C.-C. Jia, B. Zhang, Y. Zou, H.-P. Li, B.-S. Fu, Overexpression of zinc finger protein 687 enhances tumorigenic capability and promotes recurrence of hepatocellular carcinoma, Oncogenesis 6 (2017) e363.

[33] L. Yu, F. Ye, Y.-Y. Li, Y.-Z. Zhan, Y. Liu, H.-M. Yan, Y. Fang, Y.-W. Xie, F.-J. Zhang, L.-H. Chen, Y. Ding, K.-L. Chen, Histone methyltransferase SETDB1 promotes colorectal cancer proliferation through the STAT1-CCND1/CDK6 axis, Carcinogenesis (2019).

[34] L. Liu, L. Qi, T. Knifley, D.W. Piecoro, P. Rychahou, J. Liu, M.I. Mitov, J. Martin, C. Wang, J. Wu, H.L. Weiss, D.A. Butterfield, B.M. Evers, K.L. O'Connor, M. Chen, S100A4 alters metabolism and promotes invasion of lung cancer cells by up-regulating mitochondrial complex i protein NDUFS2, J. Biol. Chem. 294 (2019) 7516–7527.

[35] L. Chen, R. Liu, Z.P. Liu, M. Li, K. Aihara, Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers, Sci. Rep. 2 (2012) 18–20.

[36] K.Y. Ma, A.A. Schonnesen, A. Brock, C. Van Den Berg, S.G. Eckhardt, Z. Liu, N. Jiang, Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response-related genes, JCI Insight 4 (2019).

[37] P.T. Cagle, T.C. Allen, S. Dacic, M.B. Beasley, A.C. Borczuk, L.R. Chirieac, R. Laucirica, J.Y. Ro, K.M. Kerr, Revolution in lung cancer: new challenges for the surgical pathologist, Arch. Pathol. Lab Med. 135 (2011) 110–116.