ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

Lung Adenocarcinoma Subtyping and Feature Identification Based on Multi-omics Data Analysis

Wenzong Lu*, Hang Ma

Department of Biomedical Engineering, College of Electronic and Information Engineering, Xi'an Technological University, Xi'an, Shaanxi Province, 710021, China.
**Corresponding Author: wenzonglu@163.com*

Abstract: *Lung cancer is one of the cancers that seriously threaten human life, especially lung adenocarcinoma. Although various diagnostic and therapeutic methods continue to appear, the clinical outcomes are still gloomy. The gene coding and non-coding ribonucleic acid data of lung adenocarcinoma was analyzed by using the non-negative matrix factorization clustering, differential expression and survival analysis. Functional and pathway enrichment analysis was performed for each subtype. The computed tomography imaging features of lung adenocarcinoma were extracted and associated with weighted gene co-expression network analysis. All patients with lung adenocarcinoma could be significantly classified into three subtypes according to gene expression feature. The differential expressed genes were significantly enriched in biological processes including lung development, morphogenesis of an epithelium, receptor-mediated endocytosis for subtype 1, subtype 2, subtype 3, respectively. The number of special molecules is 340, 71, 109 for subtype 1, subtype 2, subtype 3, respectively. Some of them such as neuroendocrine convertase 1 for subtype 3, has_mir_10b for subtype 2, parathyroid hormone-related protein for subtype 1 and so on, were strongly associated with the integrate density feature in region of interest of computed tomography imaging. Subtypes of patients with lung adenocarcinoma could be accurately stratified by investigating the multi-omics data features, and several computed tomography imaging features are closely associated with the special molecules of each subtype, which is useful for selecting diagnosis and therapeutic methods.*

Keywords: *Multi-omics; Lung cancer; Molecular subtype; Weighted gene co-expression network analysis*

1. Introduction

With an estimated 2.2 million (approximately 11.4%) new cancer diagnosed cases and 1.8 million (approximately 18.0%) deaths, lung cancer is the second most frequently diagnosed cancer and the major cause of cancer-related death in 2020 year [1]. About 85% lung cancers are divided into non–small cell lung cancer such as lung adenocarcinoma, squamous cell carcinoma and large cell carcinoma histologic subtypes, and the other 15% as small cell lung cancer [2]. Lung adenocarcinoma is the most common type of lung cancer, which is further classified into several subtypes and variants by distinct cellular and molecular features. The recent development of sequencing technology has enabled us to carry out large-scale detection by RNA sequencing, whole-exon sequencing or whole-genome sequencing, which have identified significantly altered genes in lung adenocarcinoma. Several named driver genes serve as an essential role in oncogenic activity among those genes. Molecular-targeted therapy using particular inhibitors for driver gene aberrations has shown positive efficacy [3]. Advanced therapies such as selective inhibitor of epidermal growth factor receptor's tyrosine kinase domain (EGFR-TKIS) and programmed death-ligand 1 (PD-L1) have been applied to clinical treatment for lung cancer in recent years [4, 5].

MicroRNAs are also served as promising targets for treatment of various cancers, and miRNA-based lung cancer therapy has been performed in many emerging studies. Some potential miRNAs work as oncogenic and/or metastasis-promoting role such as miR-125b, miR-21, miR-135b, and some play tumor suppressor role such as miRNAs Let-7 family, miR133b, miR-181 [6]. Compared to the miRNAs, only a small proportion of lncRNAs has been uncovered to be biologically relevant although accumulating evidence indicating that the majority of them are likely to be functional [7]. According to recent evidences, lncRNAs are continuously reported to be involved in transcriptional regulation. The expression of lncRNAs is misregulated in various types of cancers, such as non-small cell lung cancer

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

[8]. Although the last decade has yielded encouraging results with lung cancer screening and systemic therapies, 5-year survival rates vary from 4-17% depending on stage and regional differences. Moreover, the lung cancer has long been a disease characterized by late-stage diagnosis [9]. Imaging-based biomarkers have been evaluated for performance in early diagnosis technology of lung cancer by low-dose computed tomography as well as an ability to predict the malignancy of it [10]. However, the prolongation of survival overall is limited because of the drug resistance, radio-resistance, and the serious side effects after various treatments for lung cancer.

In this work, the subtypes and biomarkers of lung adenocarcinoma were identified by rigorous bioinformatics and statistical analyses based on the data of expression of mRNA, miRNA and lncRNA. Imaging biomarkers were also identified for lung adenocarcinoma in computed tomography with biological interpretation by associating imaging features and biological molecule modules.

2. Materials and Methods

2.1 Data source and preprocessing

The expression of mRNA and miRNA profiles of lung adenocarcinoma patients were downloaded from The Cancer Genome Atlas (TCGA) data portal (https://tcga-data.nci.nih.gov/tcga/) by the TCGAbiolinks package of R language. The data of lncRNA and clinical trait were downloaded from TCGA data portal by SangerBox tool. The dataset of mRNA profile contained 519 primary solid tumor samples and 58 solid tissue normal samples. The dataset of miRNA profile contained 504 primary solid tumor samples and 19 solid tissue normal samples. The dataset of lncRNA profile contained 516 primary slid tumor samples and 60 solid tissue normal samples. The number of common primary solid tumor samples is 504 and common solid tissue normal samples is 19 in these three datasets.

Computed tomography (CT) imaging of lung adenocarcinoma is acquired from The Cancer Imaging Archive (TCIA) (https://public.cancerimagingarchive.net/ncia/login.jsf), which is a medical cancer imaging archive that contains imaging corresponding to the TCGA patients. 25 patients are selected in radiogenomic analysis by the molecule expression data matching the primary solid tumor images for each patient sample.

2.2 Cluster analysis based on nonnegative matrix factorization

All analyses in this work are performed using R language unless otherwise mention. The row molecule is removed when the expression value of it is zero in more than 100 samples for these three datasets. Missing value of molecule expression is supplied by the estimation methods based on K Nearest Neighbors (KNN) algorithm with the impute package of R language (number of neighbors=10). The next molecules are selected when the coefficient of variation of molecules is beyond 1.5. Subsequently, a total of 2535 molecules of 504 primary solid tumor samples were used as the input dataset of the non-negative matrix factorization (NMF). NMF was performed with the NMF package and standard method. The negative values of the dataset were zeroed to give a non-negative matrix, and the NMF was performed for ranks 2 to 10 by the Brunet algorithm. The dispersion, silhouette and cophenetic correlation coefficient score were computed, and were directly obtained using the plot function of the NMF package for each rank. The differences of subtypes were analyzed among the clusters using sigclust package, which assesses the significance of clustering by simulation from a single null Gaussian distribution. Statistical significance was set at p-value < 0.01.

2.3 Differential expression analysis

According to the results of cluster analysis based on NMF, the tumor samples are divided into 3 subtypes which respectively conclude 60, 148, 296 tumor samples. Differential expression analysis is accomplished with multiple linear regression by the limma package of R language. It estimates the fold changes and standard errors by fitting a linear model for every molecule by lmFit and the empirical Bayes statistics implemented by eBayes (adjust method='false discovery rate'). Statistical significance was set at p-value < 0.01 and absolute value of log2-fold change (log2|FC|) > 1 for 2535 molecules of each subtype.

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

2.4 Survival analysis

The samples are divided into high expression group and low expression group according to the expression value of significantly differentially expressed genes with a median value as the cutoff. The tumor samples are divided into three subtypes (groups) according to the results of cluster analysis based on NMF. The survival time and status data are from the clinical trait data. The log-rank test compares the survival times of two groups and the p-value is calculated using the chi-squared distribution by the survival package of R language. Statistical significance was set at p-value < 0.01.

2.5 Functional and pathway enrichment analysis

The database for annotation, visualization and integrated discovery (DAVID, https://david.ncifcrf.gov/) was used to classify significantly differential expression genes of each subtype by their biological processes, cellular components, and molecular functions using gene ontology (GO) and the significant transcripts (p-value <0.05) were identified by the functional annotation clustering tool. The DAVID database was also used to carry out pathway enrichment analysis with reference from kyoto encyclopedia of genes and genomes (KEGG) database website and p-value <0.05 as a cut-off point.

2.6 Imaging feature extraction

A total number of 261 imaging features of the 25 samples were evaluated, which include first order statistics, geometry features, and textural features. The first-order statistics feature is quantified tumor intensity characteristics and calculated from the histogram of all tumor voxel intensity values. Geometry feature consists of size-based tumor volume. Detailed descriptions of the imaging features extracting are described in the Supplementary Methods. For each patient sample, the CT image is selected when it contains region of interest (ROI), and the feature values of the whole image (512×512) and the ROI are extracted respectively. Each feature is expressed as the average value exclude tumor volume. For the whole image, 11 feature algorithms were implemented in Matlab software (version 2020a) and described in the Supplementary Methods. For the ROI, also called tumor outline was drawn and 7 features (area, standard deviation, mean gray value, perimeter, integrate density, skewness, kurtosis) of it were measured using the ImageJ tool after the selected CT image was converted from 16-bit to 8-bit (Figure 1).

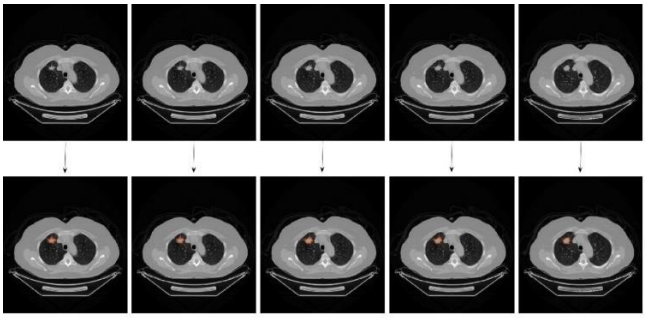

Figure 1: The figures of consecutive computed tomography slices for TCGA-50-5941 patient sample.

First row, the whole image contains region of interest. Second row, red color circle represents the region of interest.

2.7 Weighted correlation network analysis and key module identification

Weighted correlation network analysis (WGCNA)is constructed by the WGCNA package for the dataset, which includes 2535 molecules. The clinical traits dataset are the above 19 CT image features. First, the molecular expression data combined with clinical CT traits information to compute multiple soft-thresholding powers. Second, 6 is selected as the soft-thresholding power which can increase co-expression similarity to accomplish consistent scale free topology in the dataset. Third, the automatic network construction and module detection are performed with the major parameters (power = 6, TOMType = "unsigned", minModuleSize = 30, mergeCutHeight = 0.25). Finally, the correlation among modules and clinical traits is determined by correlation test, and the correlation values are displayed within a heatmap plot including p-value. Gene significance is calculated as the absolute value of the relationship between molecule expression profile and each imaging trait. Module membership is defined as the relativity of molecule expression profile and each module eigengene. The hub molecules are picked

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

out in key module when the value of module membership is more than 0.8 and the value of gene significance is more than 0.6.

2.8 Identification of special molecules for each subtype

The molecules are selected with significant differences in differential expression analysis and survival analysis for each subtype. The degree of overlap of them between subtypes was showed with venn diagram using VennDiagram package. Similarly, venn diagram also display the special feature-associated molecules for each subtype between the special molecules of subtype and the hub molecules of key module.

3. Results

3.1 Molecular subtypes identification

The molecular subtype of lung adenocarcinoma is identified based on multi-omics data (2535 molecules containing mRNA, miRNA, lncRNA) by NMF clustering algorithm in 504 samples. The values of k=3 is regarded as the optimal clustering number according to the cophenetic and silhouette values because of the first rank value for which the cophenetic and dispersion coefficient starts decreasing, and the effect of molecular clustering is better when rank value is 3 (Figure 2). The first, second, third subtype contains 60, 148, 296 different samples, respectively. The difference between these three molecular subtypes is significant ($p < 0.01$). Besides, the difference of survival analysis between subtype 1 and subtype 2 is notable ($p < 0.05$), which also appear between subtype 2 and subtype 3 (Figure 3).

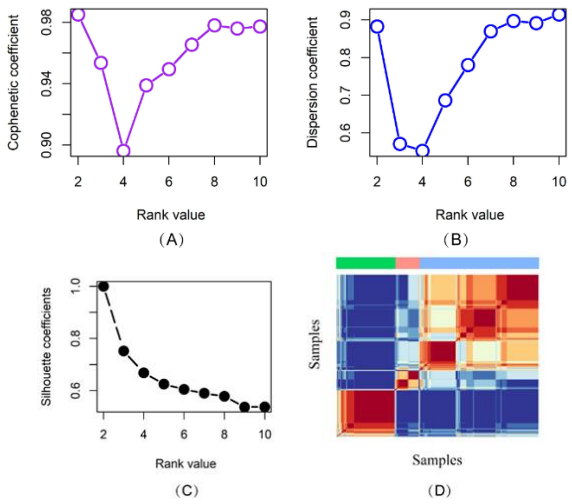

Figure 2: Cluster analysis based on nonnegative matrix factorization.

(A) cophenetic correlation coefficient at different k value. (B) dispersion correlation coefficient at different k value. (C) silhouette correlation coefficient at different k value. (D) consistency matrix of multi-omics clusters when k=3.

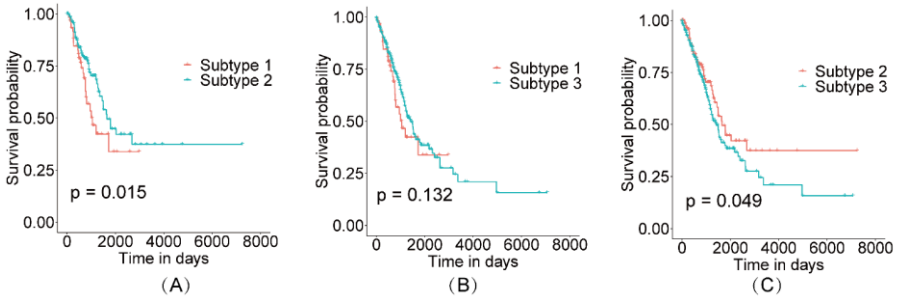

Figure 3: The difference of survival analysis among subtypes.

(A) subtype 1 versus subtype 2. (B) subtype 1 versus subtype 3. (C) subtype 2 versus subtype 3.

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

3.2 Identification of differential expression molecules

The significantly differentially expressed molecules are 694 including 387 up-regulated molecules and 307 down-regulated molecules for the subtype 1 group. The significantly differentially expressed molecules are 642 including 363 up-regulated molecules and 279 down-regulated molecules for the subtype 2 group. The significantly differentially expressed molecules are 608 including 349 up-regulated molecules and 259 down-regulated molecules for the subtype 3 group (Table 1, top 3 only).

Table 1: Analysis of molecules differentially expressed in multi-omics data (top 3 in each subtype data)

Subtype category	Ensembl ID	Gene name	logFC	P-value
Subtype 1	ENSG00000168484	SFTPC	-8.8326	4.00E-11
	hsa-mir-9-1	-	6.8588	8.03E-17
	hsa-mir-9-2	-	6.7476	1.41E-16
Subtype 2	ENSG00000168484	SFTPC	-7.5545	2.55E-12
	ENSG00000066405	CLDN18	-6.0755	3.44E-17
	ENSG00000204305	AGER	-5.8004	2.89E-22
Subtype 3	ENSG00000168484	SFTPC	-6.9661	3.62E-11
	ENSG00000066405	CLDN18	-6.4538	6.69E-17
	ENSG00000204305	AGER	-5.7795	2.82E-20

3.3 GO functional pathway enrichment analysis

It shows that the differential expressed genes are significantly enriched in biological processes including lung development, morphogenesis of an epithelium, receptor-mediated endocytosis for subtype 1, subtype 2, subtype 3, respectively. For cell component and molecular function, the different GO terms emerge among the three subtypes and the same GO terms also emerge. Although the results of KEGG signaling pathway analysis showed that the differential expressed genes were significantly enriched in complement and coagulation cascades for these subtypes, the different signaling pathways emerge among the three subtypes.

3.4 Survival analysis

The log-rank test for difference in survival gives a p-value of $p < 0.01$, indicating that it differs significantly between the high and low expression molecule groups in survival, assuming an alpha level of 0.05. These molecules include such as collagen alpha-1 (COL7A1) and protein aster-B (GRAMD1B) for subtype 1, igLON family member 5 (IGLON5) and homeobox protein Hox-D9 (HOXD9) for subtype 2, dickkopf-related protein 1 (DKK1) and protein fosB (FOSB) for subtyped 3 (Table 2, top 5 only).

Table 2: Survival analysis of multi-omics data of lung adenocarcinoma (top 3 in each subtype data)

Subtype category	Ensembl ID	Gene name	P-value
Subtype 1	ENSG00000114270	COL7A1	1.0231E-05
	ENSG00000023171	GRAMD1B	1.0696E-05
	ENSG00000104140	RHOV	1.4343E-05
Subtype 2	ENSG00000142549	IGLON5	5.6739E-04
	ENSG00000128709	HOXD9	9.8562E-04
	ENSG00000139292	LGR5	1.9400E-03
Subtype 3	ENSG00000107984	DKK1	2.6009E-06
	ENSG00000125740	FOSB	9.1589E-06
	ENSG00000163687	DNASE1L3	1.0257E-05

3.5 Construction of the CT image feature-associated molecules co-expression network

2535 molecules and 19 CT image features are used to build gene co-expression network identify the image feature-associated module hub molecules in lung adenocarcinoma. Hierarchical clustering

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

dendrogram is established based on the topology overlap matrix dissimilarity, and the dynamically cut modules with similar expression profiles are combined by the defined parameter (Figure 4A). The correlation of modules with various CT image features is calculated, such as kurtosis, skewness, volume, mean, uniformity, energy, integrate density, area, standard deviation, perimeter. The results showed that the highest positive correlation occurred between ROI integrate density and blue module ($p < 0.01$). Besides, the strongly positive correlation occurred between area or volume of ROI and cyan module ($p < 0.01$) (Figure 4B). The blue module contains 289 molecules such as parathyroid hormone-related protein (PTHRP), homeobox protein Hox-A7 (HOXA7), histamine H2 receptor (HRH2), hsa_mir_10b, hsa_mir_134 and so on.

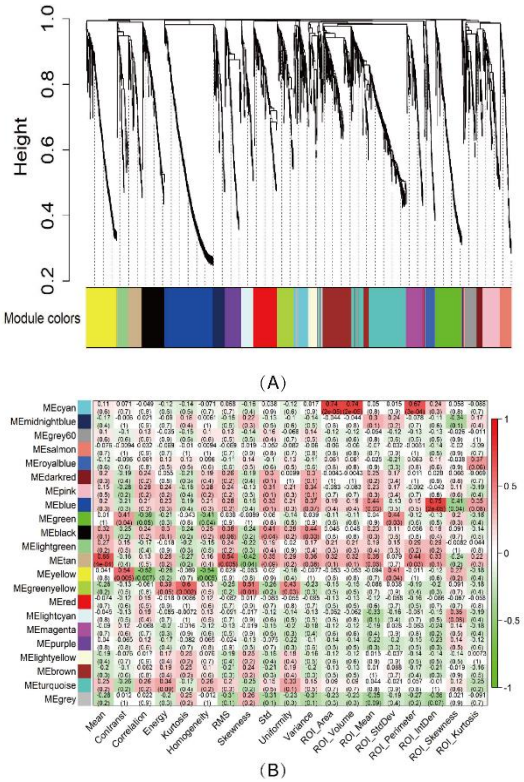

Figure 4: Construction of weighted gene co-expression network analysis.

(A) consensus gene dendrogram and module colors are showed correspondingly. (B) the module-trait relationship is displayed by correlation values and p values (p value is in parenthesis). Each row represents the named module and each column represents the traits. Red color represents positive correlation and green color represents negative correlation.

3.6 Identification of characteristic molecules of subtypes

The number of special molecules is 340, 71, 109 for subtype 1, subtype 2, subtype 3, respectivel. The characteristic molecules of subtype 1 are such as gastricsin (PGC) and hsa-mir-224. The characteristic molecules of subtype 2 are such as cytokine receptor-like factor 1 (CRLF1) and hsa-mir-34c. The characteristic molecules of subtype 3 are such as calpain-9 (CAPN9) and hsa-mir-136 (Figure 5A). In these characteristic molecules, the clinical CT imaging feature ROI integrate density is closely associated with has_mir_10b, protein AF1q (MLLT11) for subtype 2, with neuroendocrine convertase 1 (NEC1) for subtype 3, with parathyroid hormone-related protein (PTHLH) and so on for subtype 1 (Figure 5B).

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

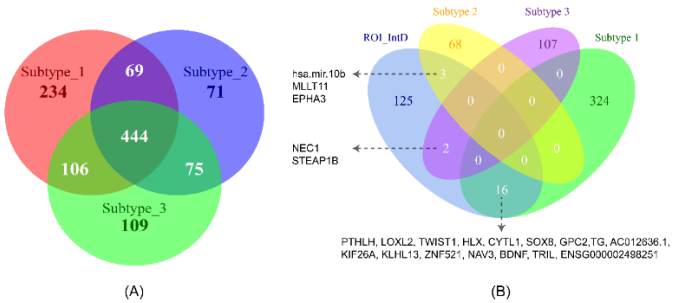

Figure 5: Identification of special molecules for the three subtypes.

(A) Venn diagram showing special molecules of each subtype. (B) Venn diagram showing the correlation of special molecules of each subtype and clinical trait such as region of interest integrate density (ROI_IntD).

4. Discussion

In this work, the molecular subtypes of lung adenocarcinoma patients are distinguished by clustering method in data mining based on the gene expression data, long non-coding RNA data, microRNA data. Patient samples with similar gene characteristics are divided into the same subtype. Our results show that the three subtypes of lung adenocarcinoma patients are identified by the multi-omics data, which are not only different in molecules expression data, but also different in their survival status. Moreover, these molecules include up-regulated and down-regulated molecules. The significant difference is displayed between these three molecular subtypes, suggesting that different subtypes are regulated by different molecules.

This study shows that the differential expressed genes are significantly enriched in biological processes including lung development, morphogenesis of an epithelium, receptor-mediated endocytosis and so on. The determination of genes that have key effects in different biological processes contributes to locating markers on specific genes, which improve the efficiency of detection and provide direction for the development of new targeted drugs. Clinical trials have also been implemented by genomic data-driven targeted treatment of lung cancer. Effectiveness-related molecular markers can also be identified through the analysis of genomic data. In recent years, some genomic detection methods have been applied to cancer prediction, early diagnosis and specific treatment in order to accurately evaluate the cancer status [11]. These methods include gene expression detection, gene mutation detection, genome sequencing, epigenetic detection and so on. The multi-omics data of cancer need to apply new technologies of machine learning and data mining to accurately classify, diagnose, test drug sensitivity and recommend treatment strategies. Some appealing progress has been obtained in this field. Study has reported that 60% of lung cancers generally occur genomic variations, which is critical to the selection of specific treatment interventions [12]. At the same time, the molecular level of tumor diagnosis gradually challenges the traditional diagnosis based on tissue section. It is showed that about 11 gene mutations occur in each cancer sample through genome sequencing technology [13]. The identification of these mutations can pick out genes related effects and may promote new therapeutic targets in the future. The relative size sorting marker of gene expression value and the proteomics research are also used to identify and obtain lung cancer markers and types.

Radiogenomics can be used to explore imaging biomarkers that can identify the genomics of cancer without the use of a biopsy in imaging genomics. Various technologies for processing high-dimensional data are used to create statistically notable correlations between computed tomography, magnetic resonance imaging and positron emission tomography imaging features and the genomics of cancer. WGCNA can assist the clinic doctor to set up a connection between clinic features and gene expressions so as to analyze both the endo-pathogenetic and exo-pathogenetic mechanisms of cancer. The radiogenomics approach has proven successful in determining the relationship between semantic image features and metagenes that represented classical molecular pathways, and it can contribute to the non-invasive identification of molecular properties of non-small cell lung cancer [14]. It has shown the potential to link imaging features of non-small cell lung nodules in CT scans to predict survival by publicly available gene expression data [15]. Similar studies in other cancers have successfully determined much of the cancer genome from non-invasive imaging features [16]. In this study, the result showed that the ROI integrate density feature of CT image is strongly associated with molecule

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

expression-based subtype in lung adenocarcinoma patients. Besides, the area and volume features of ROI also is closely related with some molecules. Several differences are generated in determining the tumor boundary when different operators manually sketch the region of interest of CT image because the boundary between most lung tumor and surrounding lung tissue is vague, which result in the bias of the extraction and collection of imaging features. Moreover, the accuracy of image feature-based prediction is related to the number of features, different extracting methods and different classifiers. Radiogenomics combines a variety of information such as image-omics, genomics and proteomics to provide extensive biological characteristics of tumor and provide a wealth of data of the precision medicine.

5. Conclusions

The three molecular subtypes are classified by multi-omics data of lung adenocarcinoma. The characteristic molecules such as NEC1 for subtype 3, has_mir_10b for subtype 2, PTHLH for subtype 1 and so on. Additionally, these molecules are strongly associated with the ROI integrate density of clinical CT imaging feature.

Acknowledgements

This research was funded by the Shaanxi Provincial Government Fund of China (2020JM-559).

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., et al. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca-a Cancer Journal for Clinicians. 71(3), 209-249.

[2] Liu, G. B., Pei, F., Yang, F. Q., Li, L. X., Amin, A. D., et al. (2017). Role of autophagy and apoptosis in non-small-cell lung cancer. International Journal of Molecular Sciences. 18(2), 367.

[3] Saito, M., Suzuki, H., Kono, K., Takenoshita, S., Kohno, T. (2018). Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. Surgery Today. 48(1), 1-8.

[4] Alsaab, H. O., Sau, S., Alzhrani, R., Tatiparti, K., Bhise, K., et al. (2017). Pd-1 and pd-l1 checkpoint signaling inhibition for cancer immunotherapy: Mechanism, combinations, and clinical outcome. Frontiers in Pharmacology. 8, 561.

[5] Gelatti, A. C. Z., Drilon, A., Santini, F. C. (2019). Optimizing the sequencing of tyrosine kinase inhibitors (tkis) in epidermal growth factor receptor (egfr) mutation-positive non-small cell lung cancer (nsclc). Lung Cancer. 137, 113-122.

[6] Zhou, Q., Sun, C. C., Li, D. J. (2017). Towards mirna based therapeutics for lung cancer. Current Pharmaceutical Design. 23(39), 5971-5972.

[7] Ma, L. N., Cao, J. B., Liu, L., Du, Q., Li, Z., et al. (2019). Lncbook: A curated knowledgebase of human long non-coding rnas (vol 47, pg d128, 2019). Nucleic Acids Research. 47(5), 2699-2699.

[8] Wang, J. L., Cai, H. B., Dai, Z. X., Wang, G. (2019). Down-regulation of lncrna xist inhibits cell proliferation via regulating mir-744/ring1 axis in non-small cell lung cancer. Clinical Science. 133(14), 1567-1579.

[9] Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., et al. (2017). Lung cancer: Current therapies and new targeted treatments. Lancet. 389(10066), 299-311.

[10] Robles, A. I., Harris, C. C. (2017). Integration of multiple "omic" biomarkers: A precision medicine strategy for lung cancer. Lung Cancer. 107, 50-58.

[11] Concolino, P., Capoluongo, E. (2019). Detection of brca1/2 large genomic rearrangements in breast and ovarian cancer patients: An overview of the current methods. Expert Review of Molecular Diagnostics. 19(9), 795-802.

[12] Devarakonda, S., Morgensztern, D., Govindan, R. (2015). Genomic alterations in lung adenocarcinoma. Lancet Oncology. 16(7), E342-E351.

[13] Vigneswaran, J., Tan, Y. H. C., Murgu, S. D., Won, B. M., Patton, K. A., et al. (2016). Comprehensive genetic testing identifies targetable genomic alterations in most patients with non-small cell lung cancer, specifically adenocarcinoma, single institute investigation. Oncotarget. 7(14), 18876-18886.

[14] Zhou, M., Leung, A., Echegaray, S., Gentles, A., Shrager, J. B., et al. (2018). Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. Radiology. 286(1), 307-315.

ISSN 2618-1584 Vol. 4, Issue 5: 1-9, DOI: 10.25236/FMSR.2022.040501

[15] Gevaert, O., Xu, J. J., Hoang, C. D., Leung, A. N., Xu, Y., et al. (2012). Non-small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data-methods and preliminary results. Radiology. 264(2), 387-396.
[16] Badic, B., Hatt, M., Durand, S., Le Jossic-Corcos, C., Simon, B., et al. (2019). Radiogenomics-based cancer prognosis in colorectal cancer. Scientific Reports. 9, 9743.