Brief Report

Transcriptome Analysis of Lung Cancer Staging

Felipe Valentim *, Carlyle Lima , Nicolas Carels

Posted Date: 30 October 2025

doi: 10.20944/preprints202510.2391.v1

Keywords: transcriptome; lung cancer; staging; progression

Brief Report

Transcriptome Analysis of Lung Cancer Staging

Felipe Leal Valentim [1,*], Carlyle Lima [2] and Nicolas Carels [2]

[1] Department of Mechatronics and Mechanical Systems Engineering of the Polytechnic School of the University of São Paulo

[2] Laboratory of Biological Systems Modeling, Center of Technological Development in Health (CDTS), Oswaldo Cruz Foundation (Fiocruz), 21040-900, Rio de Janeiro, Brazil

* Correspondence: felipelealvalentim@usp.br

Abstract

In this report, we analyzed how the profile of gene over-expression is affected by staging in lung squamous cell carcinoma. We found 23, 62 and 169 genes differentially expressed in stages I, II and III. We also validaded previously reported biomarkers for lung cancer diagnosis (ERCC1, AURKA, TPX2, BIRC5, MET, KLK10, TOP2A, PCNA, KRAS, CCNB1, CEP55, TP53, EGFR, CHEK1, CCNB2, RRM1, CDK1, MCC, UBE2C, AURKB, EXT1, PYCR1). All together, this short report points to the fact lung cancer transcriptome is a potential source for studying tumor staging and progression.

Keywords: transcriptome**;** lung cancer**;** staging**;** progression

Background

Lung cancer is the leading cause of cancer with an estimated 1.8 million deaths and 2.2 million new lung cancer cases worldwide in 2020 (Li et al., 2023). The main cause of lung cancer may be smoking (O'Keeffe et al., 2018), but studies on genetic predisposition have also found associations with race and gender (Wang et al., 2017). Lung cancer is heterogeneous and fatal, with non-small cell lung cancer (NSCLC) as its main pathological subtype. Lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) are the primary subtypes of NSCLC (Chen et al., 2014) LUSC accounts for 20%–30% of every type of lung cancer. The incidence of LUSC has decreased since the 1990s due to the efforts of smoking cessation programs (Barta et al., 2019). Specific treatments for LUSCs are still lacking, leaving advanced-stage patients with few treatment options (Lau et al., 2022) and clinical outcome remain unsatisfactory (Relli et al., 2019).

Biomarkers can be obtained through many technologies (Zhou et al., 2024). More recently, the over-expression of several genes: (i) *FGG*, *C3*, *FGA*, *JUN*, *CST3*, *CPSF4*, and *HIST1H2BH* (Wu et al., 2023) and (ii) *CCL1*, *KLRC3*, *KLRC4*, *CCL23*, and *KLRC1* (Li et al., 2021) were proposed as risk and prognostic biomarkers for LUSC, while that of other *ARC*, *CLVS2*, *ENPP5*, *FAM83D*, *HPRT1*, *HSP8*, *ITGA2*, *LCLAT1*, *LONRF3*, *MBNL2*, *MED12L*, *NACC2*, *SLC6A8*, *THBS1*, and *ZBTB4* proposed as markers of patient survival (Liu et al., 2022). However, critical clinical trial of these markers was not evaluated yet. Several ncRNA were also investigated as potential biomarkers (Wang et al., 2019; Liu et al., 2022), which show that the area is evolving quickly and that new developments could be available in a close future.

Methods

RNA-seq from TCGA samples of LUSC were downloaded from the GDC portal (https://portal.gdc.cancer.gov/) accessed on 2024-04-04. Among the 486 lung samples, 45 were paired samples (45 tumor and 45 non-tumoral samples, each from a same patient), and the remaining (441) were non-paired, which means that no control from healthy lung was available for them. The clinical sheet informed that (i) for stage I, 198 LUSC samples were non-paired while 24 were paired, (ii) for

 doi:10.20944/preprints202510.2391.v1

stage II, 130 LUSC samples were non-paired, while 17 were paired, (iii) and for stage III, 68 LUSC samples were non-paired, while 4 were paired.

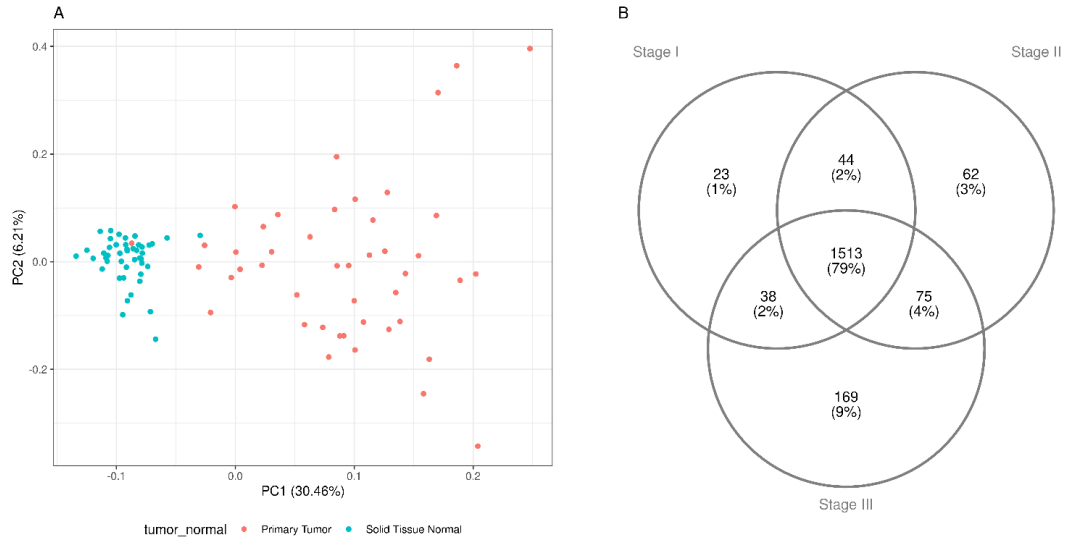
RNA-seq counts were normalized according to the reads per kilobase per million mapped reads (RPKM) methodology as described by (Mortazavi et al., 2008). Genes with an average RPK ≤ 4 were desconsidered due to the noise they may introduce.

Then, a two-phase method was used to identify stage expression signatures among LUSC genes. First, tumor genes were identified by comparing the tumor with normal samples. In this process, we identified malignant genes whose average expression significantly differed among stages by comparing 486 tumor samples by reference to the 45 normal samples, i.e, (i) for stage I, we compared 222 tumor samples with the 45 normal ones, (ii) for stage II, we compared 147 tumor samples with the 45 normal ones, and (iii) for stage III, we compared 72 tumor samples with the 45 normal ones. We identified these genes whose average expression significantly differed among stages with the abbreviation DEGS.

The average expressions in normal and tumor samples were used to calculate the genes that were significantly more expressed in tumors than in normal tissues by computing the false discovery rate (FDR) of paired t-test. From the list of significantly up-regulated genes, we considered those with a $\log_2$ fold change ≥ 1 as up-regulated. $\log_2$ fold change was calculated as the $\log_2$(average expression of tumor samples of stage i /average expression of normal samples) for $i \in \{1, 2, 3\}$. This threshold was chosen because it is consensus in the literature and because it can be detected by RT-qPCR. Note that the average expression of a given gene in tumors of a given stage might be different from that of tumors of another stage considering the same gene. However the differences were tiny, in most cases, but the average might mask larger differential expression because of the large variance among samples.

Results

We identified 4,882 genes as up-regulated among tumors of the three stages. From these, 1618, 1694 and 1795 overlapping genews were associated with stages I, II and III, respectively. The PCA (Figure 1A) shows that the first component (PC1) explained 30.46% of total variance in associated with the difference of gene expression between paired normal and tumor samples. Normal samples are arranged in a more concise group while tumor samples show more disperse pattern, both in the first and second components. The number of stage-specific genes were 23, 62, 169 for stages I, II and III, respectively (Figure 1B).



 doi:10.20944/preprints202510.2391.v1

Figure 1. RNA-seq sample composition and gene expression heterogeneity. A: PCA of tumor genes contrasting paired tumor and normal samples. B: Venn diagram of malignant up-regulated genes common to every stages (center) and DEGS (periphery).

We searched the 23, 62, 169 stage-specifgic genes in pubmed records From the 254 genes combined, we found a list of 67 genes (Yang et al., 2024; April et al., 2024; Mauricio et al., 2014; Jai et al., 2014; Pamela et al., 2017; Chuantao et al., 2021; Fangwei et al., 2022; Kaier et al., 202). Among these, 27 (40%) were present in the malignant up-regulated genes of this study.

Because we could validate 27 known, we present the list of stage I in Table 2 and the complete list for stages I, II and III in supplemental Table S1. Among the stage-specific genes of stage I (Table 1) *OLFM4, GRB7, DHX36* are related to *"stress granule"* and already known to *to be associated to lung cancer through this function*. Moreover, the FABP7 is annoted to the KEGG PPAR-signaling pathway given that PARγ in myeloid cells is known to promote lung cancer progression and metastasis (Jiyun et al., 2024).

Table 1. Known lung cancer biomarkers identified in our Tumor/Control comparisson.

		Normal		Tumor			
#	GS[1]	Av[2]	SD[3]	AV	SD	FC	Citation
1	BRCA1	2,39	0,96	10,17	6,28	4,26	Burotto et. al., 2014
2	ERCC1	22,36	3,72	32,01	12,6	1,43	Scott & Salgia, 2008; Burotto et. al., 2014
3	FGFR2	30,91	14,78	49,14	49,37	1,59	Patel et al., 2015; Villalobos & Wistuba, 2017
4	AURKA	4,26	1,89	43,31	22,78	10,17	Yunchu et. al., 2024
5	TPX2	5	4,39	127,85	87,53	25,58	Wang & Li., 2022
6	BIRC5	3,22	3,25	85,75	48,09	26,65	Yunchu et. al., 2024
7	PIK3CA	6,47	2,06	11,97	8,38	1,85	Patel et al., 2015; Villalobos & Wistuba, 2017; Zhang et. al., 2021
8	KLK10	10,62	4,75	27,39	46,8	2,58	Zhang et. al., 2021
9	TOP2A	5,64	5,9	116,07	92,53	20,59	Yunchu et. al., 2024
10	PCNA	102,48	28,78	404,4	200,94	3,95	Yunchu et. al., 2024
11	KRAS	21,86	5,41	37,03	33,92	1,69	Burotto et. al., 2014
12	CCNB1	7,85	5,34	101,89	51,12	12,98	Cai et. al., 2023
13	CEP55	2,8	2,43	49,17	29,62	17,55	Wang & Li., 2022
14	TP53	26,41	7,09	48,55	40,35	1,84	Scott & Salgia, 2008; Burotto et. al., 2014
15	EGFR	35,85	12,63	113,54	225,28	3,17	Scott & Salgia, 2008; Burotto et. al., 2014; Patel et al., 2015; Villalobos & Wistuba, 2017
16	CHEK1	1,85	0,96	18,72	9,62	10,11	Cai et. al., 2023
17	CCNB2	3,36	2,55	60,1	30,73	17,89	Cai et. al., 2023
18	BRAF	4,76	1,28	4,97	2,16	1,04	Patel et al., 2015; Villalobos & Wistuba, 2017
19	RRM1	27,57	7,56	87,18	49,05	3,16	Scott & Salgia, 2008; Burotto et. al., 2014
20	CDK1	5,52	3,16	61,52	33,84	11,15	Cai et. al., 2023
21	MCC	11,04	4,62	16,25	11,81	1,47	Zhang et. al., 2021
22	UBE2C	6,03	5,21	180,1	108,34	29,88	Yunchu et. al., 2024
23	TYMS	2,69	1,22	12,67	10,37	4,71	Yunchu et. al., 2024
24	AURKB	2,23	2,12	48,9	30,33	21,96	Yunchu et. al., 2024

 doi:10.20944/preprints202510.2391.v1

25	EXT1	29,82	8,98	49,45	25,92	1,66	Zhang et. al., 2021
26	PYCR1	8,91	5,79	71,7	56,09	8,05	Wang & Li., 2022
27	PDCD1	4,07	2,99	5,57	6,34	1,37	Patel et al., 2015

[1]GS: Gene symbol. [2]Avg.: Average. [3]S.D.: Standard deviation. [4]FC: Fold change.

Table 2. Stage-secific genes (Stage I).

	Tumor/normal			Stage I/normal		
SYMBOL	**log2fc**	**pvalue**	**FDR**	**log2fc**	**pvalue**	**FDR**
CRLF1	1,25	7,42E-02	0,1	1,49	4,63E-03	4,97E-03
LTF	0,47	4,09E-01	0,46	1,27	2,11E-03	2,31E-03
RAI14	1,14	3,28E-05	0	1,06	1,02E-11	1,56E-11
SERTAD4	1,01	3,56E-05	0	1,1	5,37E-16	9,63E-16
OLFM4	4,1	3,04E-03	0,01	6,63	1,15E-02	1,21E-02
CCDC91	0,87	2,35E-04	0	1,04	1,78E-04	2,04E-04
NPL	0,59	3,98E-02	0,06	1,01	5,30E-06	6,48E-06
GRB7	0,88	4,08E-06	0	1,34	1,56E-02	1,63E-02
NSDHL	1,08	5,75E-08	0	1,01	1,87E-39	1,05E-38
TKFC	0,91	4,21E-10	0	1,02	1,36E-40	8,06E-40
NADK2	1,05	1,47E-07	0	1,03	1,29E-21	2,92E-21
CXADR	1	4,47E-04	0	1,01	9,61E-17	1,78E-16
FABP7	9,76	3,13E-01	0,36	7,83	4,06E-03	4,37E-03
RPUSD2	0,92	3,69E-09	0	1,02	2,00E-42	1,36E-41
KRT4	0,93	2,98E-01	0,35	1,59	1,69E-02	1,76E-02
SNCG	0,42	5,30E-01	0,58	1,19	3,96E-04	4,47E-04
DHX36	1,05	1,08E-08	0	1,08	5,36E-46	4,63E-45
COPB2	0,99	7,04E-09	0	1,05	2,90E-48	3,05E-47
KPNA4	1,01	6,32E-10	0	1,01	4,12E-41	2,53E-40
AMTN	8,69	2,79E-02	0,04	7,97	1,50E-03	1,64E-03
IGHA2	1	8,76E-02	0,12	1,42	1,02E-04	1,18E-04
PCP4L1	0,74	2,51E-01	0,3	1,18	1,52E-05	1,82E-05
MARCKS	1,14	1,10E-10	0	1,03	2,63E-27	7,60E-27

Conclusions

The directive of World Health Organization (WHO) is to support the development of fast methods of lead screening against cancer, and the classification of tumors for suitable therapies. By 2020, the cancer of lung (2.21 million cases) came in second place worldwide behind that of breast (2.26 million cases), but before those of colon and rectum (1.93 million cases), prostate (1.41 million cases), skin (non-melanoma, 1.20 million cases), and stomach (1.09 million cases). Here we show that transcritpme analysis of tumor/control samples identify known biomarkers, and moreover, stage-specific genes of strage I can be associated to early stage tumor. We then prose more exploration of the GDC Cancer portal for studying tumor staging and progression.

Supplmentary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org

 doi:10.20944/preprints202510.2391.v1

Authors' contributions: Felipe Leal Valentim (FLV) and Nicolas Carels (NC) conveived the Project. FLV implemented the analysis and wrote the results. Carlyle Lima (CL) participated in discussoins.

Funding declaration: FLV had the post-doc funded by Brazlian funding agency CNPq.

Consent for publication: I hereby provide consent for the publication of the manuscript "Transcriptome analysis of lung cancer staging", including any accompanying images or data contained within the manuscript.

Ethics approval and consent to participate: This research used publicly available, non-identifiable, transcriptome of human data, already published on the GDC portal (https://portal.gdc.cancer.gov/) accessed on 2024-04-04.

Availability of data and materials: The data that support the findings of this study are openly available in GDC portal.

Acknowledgment: I thanks to Carlos Morels for the opportunity of the post-doc in Fiocruz.

Declaration of interest statement: The author declare he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Barta JA, Powell CA, Wisnivesky JP. Global Epidemiology of Lung Cancer. Ann Glob Health. 2019 Jan 22;85(1):8. doi: 10.5334/aogh.2419.
2. Mauricio Burotto,Anish Thomas,Deepa Subramaniam,Giuseppe Giaccone and Arun Rajan. Biomarkers in Early-Stage Non-Small Cell Lung Cancer: Current Concepts and Future Directions. J Thorac Oncol. 2014 Nov; 9(11): 1609–1617.
3. Kaier Cai, Zhilong Xie, Yingao Liu, Junfeng Wu, Hao Song, Wang Liu, Xinyi Wang, Yinghuan Xiong, Siyuan Gan, and Yanqin. Identification of Potential Key Genes and Prognostic Biomarkers of Lung Cancer Based on Bioinformatics. Biomed Res Int. 2023; 2023: 2152432
4. Qi Huang, Haiming Chen, Dandan Yin, Jie Wang, Shaodong Wang, Feng Yang, Jiawei Li, Teng Mu, Jilun Li, Jia Zhao, Rong Yin, Wei Li, Mantang Qiu, Erbao Zhang & Xiangnan Li. Multi-omics analysis reveals NNMT as a master metabolic regulator of metastasis in esophageal squamous cell carcinoma. npj Precision Oncology volume 8, Article number: 24 (2024)
5. Chuantao Zhang, Man Jiang, Na Zhou, Helei Hou, Tianjun Li, Hongsheng Yu, Yuan-De Tan, and Xiaochun Zhang. Use tumor suppressor genes as biomarkers for diagnosis of non-small cell lung cancer. Sci Rep. 2021; 11: 3596.
6. Fangwei Wang,Qisheng Su, and Chaoqian Li. Identidication of novel biomarkers in non-small cell lung cancer using machine learning. Sci Rep. 2022; 12: 16693.
7. Jai N. Patel, Jennifer L. Ersek, and Edward S. Kim. Lung cancer biomarkers, targeted therapies and clinical assays. Transl Lung Cancer Res. 2015 Oct; 4(5): 503–514.
8. Jiyun Zhang, Miru Tang, andJinsai Shang. PPARγ Modulators in Lung Cancer: Molecular Mechanisms, Clinical Prospects, and Challenges. Biomolecules 2024, 14(2), 190
9. Kaier Cai, Zhilong Xie, Yingao Liu, Junfeng Wu, Hao Song, Wang Liu, Xinyi Wang, Yinghuan Xiong, Siyuan Gan, and Yanqin. Identification of Potential Key Genes and Prognostic Biomarkers of Lung Cancer Based on Bioinformatics. Biomed Res Int. 2023; 2023: 2152432.
10. Lau SCM, Pan Y, Velcheti V, Wong KK. Squamous cell lung cancer: Current landscape and future therapeutic options. Cancer Cell. 2022 Nov 14;40(11):1279-1293. doi: 10.1016/j.ccell.2022.09.018.
11. Li C, Lei S, Ding L, Xu Y, Wu X, Wang H, Zhang Z, Gao T, Zhang Y, Li L. Global burden and trends of lung cancer incidence and mortality. Chin Med J (Engl). 2023 Jul 5;136(13):1583-1590. doi: 10.1097/CM9.0000000000002529.
12. Li N, Li Y, Zheng P, Zhan X. Cancer Stemness-Based Prognostic Immune-Related Gene Signatures in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma. Front Endocrinol (Lausanne). 2021 Oct 21;12:755805. doi: 10.3389/fendo.2021.755805.
13. Liu H, Li T, Dong C, Lyu J. Identification of miRNA signature for predicting the prognostic biomarker of squamous cell lung carcinoma. PloS One (2022) 17(3):e0264645. doi: 10.1371/journal.pone.0264645.

 doi:10.20944/preprints202510.2391.v1

14. Mauricio Burotto,Anish Thomas,Deepa Subramaniam,Giuseppe Giaccone and Arun Rajan. Biomarkers in Early-Stage Non-Small Cell Lung Cancer: Current Concepts and Future Directions. J Thorac Oncol. 2014 Nov; 9(11): 1609–1617.
15. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5: 621–628. pmid:18516045
16. O'Keeffe LM, Taylor G, Huxley RR, et al. Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis. BMJ Open 2018;8:e021611. doi:10.1136/bmjopen-2018-021611
17. Pamela Villalobos, Ignacio I. Wistuba. Lung Cancer Biomarkers. Hematol Oncol Clin North Am. 2017 Feb; 31(1): 13–29.
18. Jai N. Patel, Jennifer L. Ersek, and Edward S. Kim. Lung cancer biomarkers, targeted therapies and clinical assays. Transl Lung Cancer Res. 2015 Oct; 4(5): 503–514.
19. Jai N. Patel, Jennifer L. Ersek, and Edward S. Kim. Lung cancer biomarkers, targeted therapies and clinical assays. Transl Lung Cancer Res. 2015 Oct; 4(5): 503–514.
20. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of non-small cell lung cancer. Trends Mol Med (2019) 25(7):585–94. doi: 10.1016/j.molmed.2019.04.012.
21. Scott, A.; Salgia, R. Biomarkers in lung cancer: from early detection to novel therapeutics and decision making. Biomark Med. 2008 Dec 1; 2(6): 577–586.
22. Pamela Villalobos, Ignacio I. Wistuba. Lung Cancer Biomarkers. Hematol Oncol Clin North Am. 2017 Feb; 31(1): 13–29.
23. Wang, J., Liu, Q., Yuan, S. et al. Genetic predisposition to lung cancer: comprehensive literature integration, meta-analysis, and multiple evidence assessment of candidate-gene association studies. Sci Rep 7, 8371 (2017). https://doi.org/10.1038/s41598-017-07737-0.
24. Wang X, Su R, Guo Q, Liu J, Ruan B, Wang G. Competing endogenous RNA (ceRNA) hypothetic model based on comprehensive analysis of long non-coding RNA expression in lung adenocarcinoma. PeerJ (2019) 7:e8024. doi: 10.7717/peerj.8024.
25. Wu R, Ma R, Duan X, Zhang J, Li K, Yu L, Zhang M, Liu P, Wang C. Identification of specific prognostic markers for lung squamous cell carcinoma based on tumor progression, immune infiltration, and stem index. Front Immunol. 2023 Sep 29;14:1236444. doi: 10.3389/fimmu.2023.1236444.
26. Yang Yunchu, Akihiko Miyanaga,corresponding author Kuniko Matsuda, Koichiro Kamio, and Masahiro Seike. Exploring effective biomarkers and potential immune related gene in small cell lung cancer. Sci Rep. 2024; 14: 7604.
27. Yang Yunchu, Akihiko Miyanaga,corresponding author Kuniko Matsuda, Koichiro Kamio, and Masahiro Seike. Exploring effective biomarkers and potential immune related gene in small cell lung cancer. Sci Rep. 2024; 14: 7604.
28. Chuantao Zhang, Man Jiang, Na Zhou, Helei Hou, Tianjun Li, Hongsheng Yu, Yuan-De Tan, and Xiaochun Zhang. Use tumor suppressor genes as biomarkers for diagnosis of non-small cell lung cancer. Sci Rep. 2021; 11: 3596.
29. Zhou, Y., Tao, L., Qiu, J. et al. Tumor biomarkers for diagnosis, prognosis and targeted therapy. Sig Transduct Target Ther 9, 132 (2024). https://doi.org/10.1038/s41392-024-01823-2.