Genes & Diseases (2023) 10, 396–398

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

RAPID COMMUNICATION

A focused transcriptomic analysis of the TP53-regulated genes identifies the GPI-anchored molecule-like protein (GML) as a favorable prognostic predictor of lung cancer

Lung cancer is the most common cause of cancer-related mortality worldwide.[1] The current clinical staging systems cannot adequately predict the prognosis of patients with lung cancer, making it difficult to individualize the clinical treatment of the disease, resulting in poorer outcomes.

Gene expression microarrays measure the expression of thousands of genes at the same time. In the past few years, many microarray-related studies have identified gene signatures that are able to predict the prognosis of various diseases.[2] Although some of these have provided new strategies for clinical practice, there are currently no validated reports of signatures for lung cancer.

The tumor suppressor TP53 plays an important role in cancer progression. The p53 protein (encoded by *TP53*) is a transcription factor and regulates many downstream genes and important pathways involved in DNA damage repair, cell cycle arrest, and apoptosis.[3] Studies has shown the associations between *TP53* mutations and the prognosis of several types of cancer.[4] It was shown that the expression level of p53 protein was associated with the progression-free and overall survival in lung cancer.[5] However, the association between the expression of p53 target genes and the prognosis of lung cancer remains to be studied.

In this study, we analyzed the expression profiles of p53 target genes using a published microarray dataset, which included both the gene expression levels and clinical data, and is the largest reported dataset of lung cancer so far. The associations of p53 target genes with various clinical variables were initially tested in the training cohort, and then validated in two additional cohorts. The p53-related risk equation was developed as potential prognostic predictors, which may be explored as novel targets for drug development after it is validated in large prospective trials.

The gene expression data and clinical data of total 443 lung adenocarcinoma samples were collected from four institutions in the USA (Fig. 1A). The samples from the University of Michigan Cancer Center (MICC; $n = 178$) and the Moffitt Cancer Center (HLM; $n = 79$) were merged into a training cohort (Fig. S1; Fig. 1B). The samples from the Dana-Farber Cancer Institute (DFCI; $n = 82$) and the Memorial Sloan-Kettering Cancer Center (MSKCC; $n = 104$) were used as two independent validation cohorts. The distributions of clinical characteristics, such as age, gender and stage, were similar among the four institutions except for stage in the DFCI dataset. Thus, the samples from the four institutes were considered comparable.

In the microarray platform, several probes may reflect the expression of a single gene. In this microarray data, the *MDM2*, *SH2D1A*, and *VDR* genes each had four probes, and *BTG2* and *IGFBP3* each had two probes, while all other genes had one probe (Table S1). The expression level evaluated for risk stratification was the average expression of all probes for the genes that had more than one probe in the training cohort; and the expression levels ranged from 3.56 to 9.88 on a logarithmic scale (Fig. 1B). These findings indicated that the array experiments were acceptable for further analysis.

We chose 13 of the 15 p53-reguated genes to be included in the analysis (Table S1). A multivariate Cox proportional hazards regression with covariates of age, gender, and stage was used to evaluate the associations between genes and overall survival in the training cohort. The results showed that only GPI-anchored molecule-like protein (GML) was

Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.08.022

A

	MICC	LM	DFCI	MSKCC
Sample size	178	79	82	104
Average age	64	67	62	65
Gender-male	57%	51%	56%	36%
Stage-I (%)	65%	54%	68%	61%
Stage-II (%)	17%	26%	32%	19%
Stage-III (%)	18%	19%	0%	20%

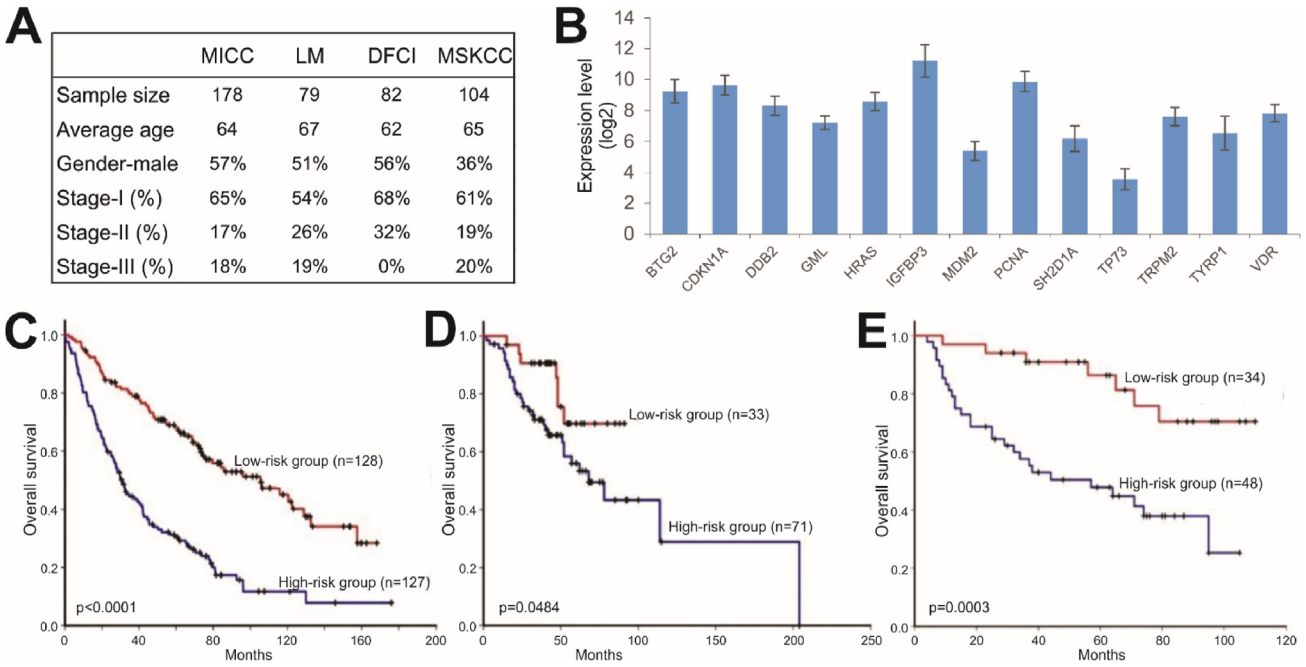

Figure 1 Results of data analysis. **(A)** Clinical characteristics of the 422 lung adenocarcinoma patients in the four cohorts. **(B)** Average expression levels of the 13 p53 target genes in the training cohort (UICC plus HLM, $n = 257$). **(C)** Survival analysis and risk score in the training cohort. **(D)** Survival analysis and risk score of GML in the first independent validation cohort (MSKCC, $n = 104$). **(E)** Survival analysis and risk score of GML in the second independent validation cohort (DFCI, $n = 82$). DFCI, Dana-Farber Cancer Institute; HLM, H. Lee Moffitt Cancer Center; MICC, University of Michigan Cancer Center; and MSKCC, the Memorial Sloan-Kettering Cancer Center.

significantly associated with the overall survival. Hence, the clinical risk score was generated: Risk = 0.02337 × (age + 0.22242) × (sex + 0.76568) × (stage + 0.48335) × *GML*.

The coefficients for each covariate were estimated from a multivariate Cox proportional hazards regression. The risk score for each sample can be calculated using the above formula. The median of the clinical risk score was set as the cut-off point in the training cohort. The samples with a risk score higher or equal to the median were considered high-risk patients and those with a lower score were considered low-risk patients. The results showed that patients with high risk had a shorter overall survival ($P < 0.0001$, log rank test) (Fig. 1C).

We next validated the GML risk score in two independent cohorts. The same coefficients and cut-off point were not re-estimated, and were directly applied to the two independent cohorts because it was considered that the overall characteristics were comparable among all population groups in this study. The results showed that the patients with low GML risk scores had a longer overall survival than those with high risk GML scores in the MSKCC cohort ($P = 0.0484$, log rank test) (Fig. 1D). Similar results were also found in the DFCI cohort ($P = 0.0003$, log rank test) (Fig. 1E).

In this study, we examined the expression levels of p53 target genes in a large microarray dataset. Then, the associations of p53 target genes with the overall survival were evaluated using multivariate Cox proportional hazards regression analyses. We only found GML to be significantly associated with the overall survival in the training cohort. Patients were subsequently defined as high- or low-risk based on the GML risk score. The results were validated in two additional unique cohorts. It is noteworthy that the parameters used to calculate the GML risk score were not re-estimated in the two validation cohorts.

A p53 binding site is present in the promoter or intron regions of GML. It has been reported that cancer cells with higher expression levels of wild type TP53 were sensitive to cisplatin chemotherapy. Our GML-based risk prediction equation incorporated the effects of age, gender, and stage, and has shown good performance for the identification of patient risk. The association of p53 itself with the overall survival was not significant in either the training cohort or the two validation cohorts in this study. Nonetheless, the mutation status of *TP53* in the individuals included in this study was unknown. Hence, the effects of the TP53 mutations on the GML-based risk score cannot be assessed. Future studies should validate the prognostic value of GML for lung cancer in a large cohort of multiple centers, and/or investigate the potential effects of TP53 mutations on GML-based risk score. Collectively, the GML-based risk equation may have potential to be a valuable prognostic predictor of lung cancer, and GML may serve as a new target for drug development.

Author contributions

Yutong Wang and Shanshan Wang conducted the analysis and data collection, drafted the manuscript. Yanmei Cui, Jie Zhang, Shuang Geng, Honglei Yin and Simiao Zhang provided resources and data analysis. Qiufang Li and Yunliang Wang designed the experiments, and revised/edited the manuscript. All authors read and approved the submitted manuscript.

Conflict of interests

The authors declare no conflict of interests.

Funding

The reported work was supported in part by a research grant from the Shandong Natural Science Foundation (No. ZR2019MH065).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2022.08.022.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin*. 2013;63(1):11—30.
2. Hsu YC, Chen HY, Yuan S, et al. Genome-wide analysis of three-way interplay among gene expression, cancer cell invasion and anti-cancer compound sensitivity. *BMC Med*. 2013;11:106.
3. Levine AJ, Oren M. The first 30 years of p53: growing ever more complex. *Nat Rev Cancer*. 2009;9(10):749—758.
4. Steels E, Paesmans M, Berghmans T, et al. Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis. *Eur Respir J*. 2001;18(4):705—719.
5. Mogi A, Kuwano H. TP53 mutations in nonsmall cell lung cancer. *J Biomed Biotechnol*. 2011;2011:583929.

Yutong Wang [a,1], Shanshan Wang [c,1], Yanmei Cui [e], Jie Zhang [a], Shuang Geng [c], Honglei Yin [c], Simiao Zhang [d], Qiufang Li [b], Yunliang Wang [d,*]

[a] *School of Public Health and Laboratory of Qilu Medical University, Zibo, Shandong 255300, China*
[b] *School of Nursing and Health of Zhengzhou University, Zhengzhou, Henan 450052, China*
[c] *Department of Neurology, 960th Hospital of PLA, Zibo, Shandong 255300, China*
[d] *Department of Neurology, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450014, China*
[e] *Shangqiu Medical College, Shangqiu, Henan 476100, China*

*Corresponding author.
E-mail address: wangyunliang81@163.com (Y. Wang)

10 February 2022
Available online 11 October 2022

[1] Co-first authors.