A genomic-clinicopathologic nomogram predict survival for patients with laryngeal squamous cell carcinoma

Jie Cui

Guangzhou Medical University Affiliated Cancer Hospital

Qingquan Wen

Guangzhou Medical University Affiliated Cancer Hospital

Xiaojun Tan

Guangzhou Medical University Affiliated Cancer Hospital

Jinsong Piao

Guangzhou Medical University Affiliated Cancer Hospital

Qiong Zhang

Guangzhou Medical University Affiliated Cancer Hospital

Qian Wang

Guangzhou Medical University Affiliated Cancer Hospital

Lizhen He

Guangzhou Medical University Affiliated Cancer Hospital

Yan Wang

Guangzhou Medical University Affiliated Cancer Hospital

Zhen Chen

Shunde Hospital of Southern Medical University

Genglong Liu (✉ lglong3@mail2.sysu.edu.cn)

Affiliated Cancer Hospital of Guangzhou Medical <https://orcid.org/0000-0002-9163-4396>

Research article

Keywords: laryngeal squamous cell carcinoma; lncRNAs signature; nomogram; prediction; overall survival;

Posted Date: August 15th, 2019

DOI: <https://doi.org/10.21203/rs.2.12975/v1>

License: (cc) (i) This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Disease Markers on November 20th, 2019. See the published version at <https://doi.org/10.1155/2019/5980567>.

Abstract

Long non-coding RNAs (lncRNAs), which have little or no ability to encode proteins, have attracted special attention due to their potential role in cancer disease. We aimed to establish a lncRNAs signature and a nomogram incorporating the genomic and clinicopathologic factors to improve the accuracy of survival prediction for laryngeal squamous cell carcinoma (LSCC). LSCC RNA sequencing (RNA-seq) data set and the matched clinicopathologic information were downloaded from the Cancer Genome Atlas (TCGA). Using univariable Cox regression and Least absolute shrinkage and selection operator (LASSO) analysis, we developed a thirteen lncRNAs signature related to prognosis. On the basis of multivariable Cox regression analysis results, a nomogram integrating the genomic and clinicopathologic predictors was built. The predictive accuracy and discriminative ability of the inclusive nomogram were confirmed by calibration curve and a concordance index (C-index), and compared with TNM stage system by C-index, receiver operating characteristic (ROC) analysis. Decision curve analysis (DCA) was conducted to assess clinical value of our nomogram. Consequently, thirteen overall survival (OS) -related lncRNAs were identified, and the signature consisting of the selected thirteen lncRNAs could effectively divide patients into high-risk and low-risk subgroup, with the area under curve (AUC) of 0.89 (3-year OS) and AUC of 0.885 (5-year OS). Independent factors derived from multivariable analysis to predict survival were margin status, tumor status and lncRNAs signature, which were all assembled into the nomogram. The calibration curve for the survival probability showed that the predictions based on the nomogram were in good coincide with actual observations. The C-index of the nomogram was 0.82 (0.77-0.87), and the area under curve (AUC) of nomogram in predicting overall survival (OS) was 0.938, which were significantly higher than traditional TNM stage. Decision curve analysis further demonstrated that our nomogram had larger net benefit than TNM stage. In summary, an inclusive nomogram for patients with LSCC, comprising genomic and clinicopathologic variables, generates more accurate estimations of the survival probability when compared TNM stage alone, but more additional data remains needed before being used in clinical practice.

Background

Laryngeal squamous cell carcinoma (LSCC), as an aggressive malignancy, accounts for 85–95% of all laryngeal cancer and is one of the most prevalence cancers in head and neck region [1]. According to age-adjusted morbidity estimates, in the United States there are 5.4 males per 100,000 residents and 1.1 females per 100,000 residents [2]. In Europe, these figures slightly shift, higher for men with 8.8 accidents per 100,000 inhabitants and lower for women with 0.8 accidents per 100,000 inhabitants, respectively [3]. According to the latest report, the 5-year overall survival rate is approximately 60% for treated laryngeal cancer patients [4]. Although multimodal therapies have been applied in LSCC, the survival outcome are still unsatisfactory, especially for advanced laryngeal carcinoma [5].

Currently, the existing American Joint Committee on Cancer (AJCC) tumor-node -metastasis (TNM) stage system [6], based on anatomical information, is often adopted when involving the estimated prognosis of patients. LSCC is consisted of a heterogeneous histological subtype with extensive clinical course

variations [7]. As a result, a significant proportion of patients with inaccurate stage may receive over-treatment or under-treatment. For instance, overstage might subject a patient to needless adjuvant chemoradiotherapy; conversely, understage is likely to result in recurrence or even death after surgery. Hence, identifying reliable and novel markers/models to improve accuracy of prediction in LSCC patients is very urgent and necessary to optimize the treatment planning and benefit patients.

As previous genome researches revealed, more than ninety percent of the human genome is actively transcribed into non-coding RNA (ncRNAs) [8]. Conventionally, ncRNA family is loosely classified into two categories based on molecular size: small non-coding RNA (the length less than 200nt; eg microRNA) and long ncRNA (the length more than 200nt; lncRNA) [9]. Unlike protein-coding RNAs, the expression patterns of the lncRNAs are more specific. A large number of researches have reported the diverse biological functions of lncRNAs, such as tumorigenesis, tumor progression, as well as metastasis [10]. lncRNA can be as a new cancer biomarker, which represents a large number of potential molecular drivers in human cancer disease [11]. In the past several years, lncRNAs signatures have been reported to evaluate prognostic of cancers, including head and neck cancer, cervical carcinoma and gastric cancer [12-14]. But, the lncRNA signature that is applied to predict the overall survival (OS) outcome of LSCC has not been found yet.

In present study, we hypothesized that inclusive nomogram containing genomic and clinicopathological factors can improve the prediction accuracy of survival probability. By mining the expression data of lncRNAs in The Cancer Genome Atlas (TCGA), we appraised lncRNAs that were significantly related to survival outcomes, and then developed a lncRNAs signature. An inclusive nomogram for predicting survival status was established by further integrating lncRNAs signature with clinicopathological factors. We assessed the predictive ability and clinical application of the nomogram and compared it to the TNM stage. In addition, we evaluated the prediction effect of the nomogram in clinical subgroups (advanced LSCC and early LSCC).

Methods

Collection of public data from TCGA

LSCC RNA sequencing (RNA-seq) data set and relevant clinicopathological information including the age, sex, smoke history, alcohol history, number of lymph nodes (LN), number of positive LNs, lymph node ratio(LNR), margin status, tumor status, histologic grade, T stage, N status, TNM stage, mutation count, fraction genome altered, and overall survival(OS) time were obtained from TCGA database(<https://gdc.cancer.gov/>). A total of 109 patients with complete follow-up data were extracted, which recorded before April 14, 2019. The clinical end point was overall survival (OS), defined as the time from surgery to death. In addition, patients who were alive are considered as censored cases at last follow-up.

Given that the expression level of lncRNAs is relatively low compared with non-coding RNA, it is likely that some lncRNAs have not been analyzed during the sequencing procedure of lncRNAs. Considering

this possibility, we defined lncRNAs as being expressed abundantly when its expression level is above 0 and occurs more than 50% in the total samples. The final expression level of lncRNAs was represented as $\log_2(x + 1)$ of the original expression level.

Construction and confirmation of a lncRNAs signature

First, moderated t-statistics method and Benjamini–Hochberg procedure are used to identify distinct differential lncRNAs between normal tissues and LSCC tissues, with the false discovery rate (FDR) <0.05 and $P < 0.05$ for filtration. Next, univariable Cox regression analysis is applied to pick out prognostic related lncRNAs with statistically significant ($P < 0.01$). After primary filtering, a Least Absolute Shrinkage and Selector Operation (LASSO) analysis is established to select candidate lncRNAs with penalty parameters tuning adjusted by 10 times cross validation [15], then a signature based on these well-selected lncRNAs is developed.

The risk score formula is generated by integrating these prognostic related lncRNA, weighted by their respective LASSO regression coefficients. According to this formula, each patient's risk score was calculated, and patients were classified into high-risk or low-risk group on the basis of the optimal cut-off value, which was adopted in the maximum sensitivity and specificity by using receiver operating characteristic (ROC) curve (time-independent). The survival difference between high-risk group and low-risk group are further compared by the Kaplan-Meier analysis (log-rank test). Stratified analysis based on various clinical characteristics is conducted to evaluate the discrimination ability of lncRNAs signature.

Function enrichment analysis of the prognostic lncRNAs

In TCGA dataset, according to their expression level, Pearson correlation algorithm is performed between the identified lncRNAs and the protein-coding genes (mRNAs). The $P < 0.001$ and correlation coefficient >0.4 are defined significant correlation. The potential biological functions of the lncRNAs target genes are investigated by using Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). DAVID is a common bioinformatics tool (<http://david.abcc.ncifcrf.gov/>, version 6.8) [16], which is applied to investigate the biological processes of the selected lncRNAs. Adjusted P values with a FDR < 0.05 for GO analysis and KEGG pathways are considered remarkably enriched functional annotations.

Genomic-clinicopathologic nomogram

To establish a genomic-clinicopathologic nomogram, we carried out univariate and multivariate Cox regression analysis to appraised clinical risk parameters associated with survival. Then, the lncRNAs-based signature, together with the risk parameters, were applied to develop a comprehensive nomogram.

The performance of the model was evaluated by calibration and discrimination. Discrimination is the model's ability to differentiate between patients who survived versus those who did not. The concordance index (C-index) was computed to assess the discrimination. Besides, we illustrated the discrimination by dividing the data set into three groups according to the scores generated by the nomogram. We plotted a Kaplan–Meier curve for all three groups. In addition, calibration curves with plotting the nomogram predicted probabilities against the observed outcomes were graphically evaluated.

Furthermore, we used ROC analysis to investigate and compare the discrimination ability of the nomogram with TNM stage or lncRNAs signature. Decision curve analysis (DCA) was used to assess the clinical usability and net benefit of the predictive model, and compared with traditional TNM staging or lncRNAs signature [17]. Finally, we evaluated the predictive accuracy of the comprehensive nomogram in clinical subgroups (advanced LSCC and early LSCC).

Statistical analysis

Categorical variables are provided as proportions (%). Continuous variables are described as medians (interquartile ranges [IQRs]) if the distribution was non-normal, and as means (standard deviations [SDs]) if the distribution was normal.

If there were missed values in some of the potential predictors, these missing data would be imputed, as complete case analysis would improve the statistical power and reduce potentially biased results [18]. Multiple imputation (MI) was conducted to interpolate the missing data as the missing data were deemed to randomly miss after analyzing patterns of them [19]. We used Markov chain Monte Carlo (MCMC) function to perform MI, and selected five iterations to account for possible simulation errors.

LASSO algorithm was conducted with “glmnet” packages, and ROC analysis was generated with “timeROC” and “survivalROC” packages. The nomogram and calibration plots were done with “rms” packages, and DCA was performed with the “stdca.R”.

SPSS statistics 22.0 and R software (R version 3.5.2) were used to conduct the statistical analysis. A $P < 0.05$ with two-sided would be recognized as statistically significant.

Results

Demographic parameters and OS outcome of LSCC patients

In the current study, 109 LSCC patients with available lncRNAs data and clinicopathological characteristics were included. The basic clinicopathological features of these LSCC patients were summarized in **Table 1**. The median OS time was 23.32 months (from 2.53 to 210.81 months). Of all the 109 LSCC patients, 49 patients (45%) died during follow-up and the 3-year and 5-year OS rates were 79.7% (71.7–87.7%) and 64.5% (54.7–74.3%), respectively.

Construction and confirmation of a lncRNAs signature

Based on the primary filter criteria mentioned in the methods section, we obtained a list of 1160 different lncRNAs (**Supplementary material 1**). Then, using univariable Cox regression analysis, we identified 31 prognostic related lncRNAs (**Supplementary material 2**). Final, LASSO algorithm were used to shrink and pick out the OS-related lncRNAs (**Figure 1**), which were AC007907.1, AC025419.1, AC078993.1, AC090241.2, AL158166.1, AL355974.2, AL596330.1, HOXB-AS4, KLHL6-AS1, LHX1-DT, LINC00528, LINC01436 and TTTY14, to build a lncRNAs-based signature.

In order to better investigate the value of lncRNAs signature in predicting prognosis, a risk score was established, with the coefficients weighted by LASSO Cox regression. The risk score was generated as follows: risk score = $(0.2102 * AC007907.1) + (0.0045 * AC025419.1) + (0.1377 * AC078993.1) + (-0.3675 * AC090241.2) + (-0.0652 * AL158166.1) + (0.0180 * AL355974.2) + (0.1208 * AL596330.1) + (0.0969 * HOXB-AS4) + (0.2227 * KLHL6-AS1) + (0.1541 * LHX1-DT) + (-0.0647 * LHX1-DT) + (-0.0750 * LINC01436) + (-0.1360 * TTTY14)$. Using ROC curve to generate the optimal cutoff value for the risk score, patients were categorized into high-risk group and low-risk group. As w-as shown at **Figure 2**, patients with high-risk score were more likely to die and had shorter OS time than patients with low risk score (19.74 vs 108.9 months, HR=5.79, 95%CI: 3.18-10.54, $P<0.0001$). The lncRNAs signature had a superior prediction effect, with AUC of 0.89 (3 year OS) and AUC of 0.885(5 year OS) (**Figure 2C**). Additionally, 13 lncRNAs signature in subsets of patients with different clinical characteristics were analyzed by stratification analysis. When stratified according to clinical variables (tumor size, node status, TNM stage), 13 lncRNAs signature remained a clinically and statistically significant prognostic model (**Figure S1**).

Functional prediction of the 13 lncRNAs

To explore the potential function of the 13 lncRNAs, a total of 237 protein-coding genes (mRNAs) were identified significantly correlated with at least one of the 13 lncRNAs ($P<0.001$ and Pearson coefficient >0.4), which were deem to eligible for pathway enrichment. The 13 lncRNAs were mainly related with human papillomavirus infection, focal adhesion and protein digestion and absorption (**Figure 3A**). And KEGG pathway analysis revealed that 13 lncRNAs related target genes mainly enriched in metalloendopeptidase activity, extracellular matrix structural constituent and metallopeptidase activity (**Figure 3B**).

Development of genomic-clinicopathologic nomogram predicting OS in LSCC patients

Using univariate Cox analysis, we identified four variables, including sex, margin status, tumor status and lncRNAs signature, were associated with survival probability (**Table 2**). Multivariable analysis continued to verify that margin status, tumor status and lncRNAs signature, were independent risk factors for OS. Based on multivariate analysis of OS, a genomic-clinicopathological nomogram was built to predict OS in 3 and 5 years (**Figure 4**). The C-index of the nomogram for OS prediction was 0.82 (0.77-0.87) (**Table 3**). The calibration plot of OS probabilities for 3 and 5 years showed the best consistency between the nomogram prediction and the actual observations (**Figure S2**). Additionally, Kaplan-Meier curve was performed to analyze the discrimination ability of the nomogram to predict OS, and a significant statistical difference was found among the three subgroups (Log-rank $P < 0.0001$) (**Figure S3**).

Comparison of predictive performance and clinical usefulness between nomogram and TNM stage or lncRNAs signature

To evaluate the predictive ability of nomogram, we compared nomogram model with AJCC TNM stage model and lncRNAs signature. As was shown at **Table 3**, the C-index of nomogram was higher than that of TNM stage (0.53 (0.45-0.61)) and lncRNAs signature (0.78 (0.71-0.85)). Likelihood ratio test, linear trend χ^2 test and akaike information criterion all showed that the nomogram was better than TNM stage or lncRNAs signature. ROC analysis also indicated that the nomogram (AUC 0.938) had higher prediction efficiency than TNM stage (AUC 0.533) or lncRNAs signature (AUC 0.847) (**Figure 5A**). Finally, DCA was conducted to compare the clinical usefulness of the nomogram to that of traditional TNM stage and lncRNAs signature. According to the continuity of potential death threshold (x-axis) and the net benefit of risk stratification using the model (y-axis), DCA visually revealed that the inclusive nomogram was superior to the traditional TNM staging or lncRNAs signature (**Figure 5B**).

Furthermore, ROC analysis in clinical subgroups (advanced LSCC and early LSCC) was conducted to assess the discrimination ability of the nomogram. As shown in **Figure 6**, encouragingly, the nomogram presented good discrimination ability in the advanced LSCC subgroup (AUC 0.951; **Figure 6A**) and the early LSCC subgroup (AUC 0.811; **Figure 6B**). Moreover, according to the best cutoff values, the patients in each subgroup were classified into low-risk group and high-risk group. Notably, the low-risk group was more likely to survive in two subgroups (**Figure 7A and Figure 7B**).

Discussion

Analyzing LSCC RNA sequencing (RNA-seq) data set and relevant clinical parameter of 109 LSCC patients from TCGA, we identified thirteen lncRNAs related to OS. On the basis of these lncRNAs, we developed a lncRNAs signature, which could accurately categorized patients into high-risk status and low-risk status. Additionally, we built a visually inclusive nomogram, integrating lncRNAs signature and clinicopathologic variables to predict survival in LSCC patients underwent surgery resection. The nomogram effectively predicted survival rate, with a bootstrapped corrected C-index of 0.73 and AUC of 0.938, which possessed better predictive ability and clinical usability than TNM stage alone.

Increasing number of studies have found that lncRNAs may be exploited as potential effective biomarkers in diagnosis, progression and prognosis of LSCC [20-22]. Basing a comprehensive lncRNAs profile for LSCC, Shen et al [20] identified AC026166.2-001 and RP11-169D4.1-001 as new lncRNAs with accurate diagnosis ability for LSCC, were independent factors for prognosis and may be potential therapeutic targets. A study of lncRNAs microarray by Chen et al. [21] uncovered that lncRNA AC008440.10 was significantly related to LSCC stage, lymph node metastasis (LNM) and survival time. Recently, Zhao et al. [22] confirmed that LINC00668, was up-regulated in LSCC, probably in vitro promote the malignant phenotypes of cells and the author deduced that LINC00668 may enhance the stability of RAB3B mRNA by binding its 3'UTR. Notably, He et al. [23] collected data from the open Gene Expression Omnibus(GEO), reported that 18-mRNA and one-lncRNA module were correlated with disease-free survival (DFS) of LSCC patients and it effectively divided patients into high-risk group and low-risk group with different DFS outcomes, independent of patient age and tumor grade. Similarly, Wu et al. [24], using data from GEO, constructed a potential panel of two-lncRNAs signature, including RP11-169K16.4 and RP11-107E5.3, to predict recurrence of patients with laryngeal carcinoma and confirmed that it was independent predictors of laryngeal cancer patients. These studies suggested the potential clinical implications of lncRNA in improving the prognosis prediction of LSCC. However, it should be noted that the lncRNAs signature predicting the overall survival (OS) outcome of LSCC has not been reported yet. Hence, in the current study, using TCGA database containing large-scale lncRNAs expression data, we aimed to identify OS-related lncRNAs and establish a lncRNAs signature, which may be more valuable for LSCC patients to optimize tailored treatment in the era of precision medicine.

To our knowledge, this is the first study constructed an inclusive nomogram, combining lncRNAs signature and clinicopathologic factors, for predicting survival probability in patients with LSCC. We built a lncRNAs signature, consist of AC007907.1, AC025419.1, AC078993.1, AC090241.2, AL158166.1, AL355974.2, AL596330.1, HOXB-AS4, KLHL6-AS1, LHX1-DT, LINC00528, LINC01436 and TTTY14, could effectively classified patients into high-risk group with shorter OS and low-risk group with longer OS. Using stratified analysis, lncRNAs signature shown perfect discrimination ability, regardless of tumor size, node status, and TNM stage. Additionally, we identified three independent predictors that margin status, tumor status and lncRNAs signature, which were all embedded into the nomogram. In this study, in consideration of homogeneity, and ability of discrimination and risk stratification of the model, the performance of the nomogram in predicting survival probability is superior to the TNM staging system. The advantage of the current nomogram is that it integrated genomic and clinicopathological variables, which are important for predicting survival risk, but cannot be obtained by TNM stage system. Remarkably, DCA results showed that LSCC survival-related treatment decision based on the nomogram led to more net benefit than treatment decision based on TNM stage or lncRNAs signature, or treating either all patients or none. Taken together, the present nomogram would be clinically useful for the clinicians in tailoring survival-associated treatment decision.

It is also worth to mention that an important feature of our comprehensive nomogram may be the ability to stratify clinical subgroups, including early LSCC and advanced LSCC. Patients diagnosed with early LSCC are generally considered to have a low survival risk, and therefore do not receive adjuvant treatment

after radical resection. Nevertheless, some patients in the clinically low-risk subgroup (early LSCC) are at high risk of survival, and they are likely to benefit from adjuvant treatment or intensive follow-up plan. Likewise, Patients diagnosed with advanced LSCC are usually identified as high risk survival status and need to receive adjuvant therapy underwent laryngectomy. Nevertheless, several patients in the clinically high-risk subgroup (advanced LSCC) are at low risk of survival, and they may not benefit from adjuvant therapy or intensive follow-up plan. It is an arduous challenge to accurately predict survival risk of patients. Encouragingly, our nomogram presented perfect discrimination capacity in early LSCC and advanced LSCC subgroups. Hence, our nomogram probably benefit a large proportion of the patients who might be considered at high risk of survival in early LSCC subgroup or might be considered at low risk of survival in advanced LSCC subgroup.

Among the thirteen OS-related LINC01436, and TTTY14 have been previously reported to be related with cancers, such as in non-small cell lung cancer (NSCLC), oral squamous cell carcinoma (OSCC) and gastric cancer [25-27]. LINC01436 has been revealed to be potential prognostic biomarkers for NSCLC [25]. In terms of mechanism, LINC01436, acted as a microRNA (miR)-30a-3p sponge, regulated the expression level of its target gene EPAS1. TTTY14 (testis-specific transcript, Y-linked 14) was thought to be associated with human papillomavirus (HPV) induced tumorigenesis, the expression level of TTTY14 was observedly differed between HPV inactive/negative group and HPV active group [28], in accordance with our functional enrichment analysis. In additional, TTTY14 was identified as significantly correlated with overall survival in OSCC and gastric cancer [26, 27]. Hence, further characterization of molecules such as AC007907.1, AC025419.1, AC078993.1, AC090241.2, AL158166.1, AL355974.2, AL596330.1, HOXB-AS4, KLHL6-AS1, LHX1-DT, LINC00528, LINC01436 and TTTY14 will provide new perspective for the development and progress of LSCC, and aided to find potential therapeutic targets for LSCC patients.

Consistent with previous studies, margin status was found to be significant association with survival among patients with LSCC in the present study [29, 30]. Published trials about LSCC did not reported that tumor status which was an independent risk factor for OS in our study was related to prognosis of LSCC. However, tumor status has been confirmed that it was independent prognostic factors for survival in hepatocellular carcinoma (HCC) [31]. Additionally, we identified male was positively associated with OS probability in the univariate Cox analysis, inconsistent with previous trials that male was poor prognosis for LSCC [32, 33]. Nevertheless, the effect of gender on the prognosis of OS was not statistically significant in multivariable Cox analysis. In addition to these clinical factors, as expected, the lncRNAs signature was an effective independent prognostic factor for the prediction of patients with LSCC.

Although our nomogram demonstrated impressive performance in LSCC survival prediction, there are specific limitations associated with our trial. First, the presented nomogram based only on TCGA database with limited sample sizes for LSCC, are not yet suitable for general application prior to validation of the predictive models with independent testing cohort. So external and multicenter prospective cohorts with large sample sizes are still needed to validate the clinical application of our model.

Second, Missing variables were a source of defect in this evaluation. Such as extracapsular spread [34, 35], lymphovascular invasion status [35], perineural invasion [35] and human papillomavirus (HPV) [36, 37] as important prognostic parameters for LSCC patients, weren't well recorded in TCGA database. Notably, our functional enrichment analysis found that the prognostic lncRNAs was significantly associated with human papillomavirus. Published researches and meta-analysis indicated HPV positive laryngeal cancer patients is sensitive to radiotherapy and chemotherapy and showed inferior survival [36, 37]. Hence, we recommend that future studies should added value of those factors in a multivariable prediction model to improve the accuracy of prediction in LSCC patients

Third, our selection of factors was limited to those available in our database. On account of the anonymous database, we cannot extend our database with characteristics such as race, insurance status, comorbidity, hemoglobin level, albumin, tumor hypoxia, and TP53 mutation, which were frequently reported prognostic factors of patients with LSCC [38-40]. Further efforts to incorporate more patient-specific, tumor-specific and molecular factors will potentially help to improve the performance of the present model.

Fourth, we do not explore the underlying biological function and pathways of the prognostic lncRNAs, so further studies are needed to uncover the related mechanisms.

Conclusions

We have built visually comprehensive nomogram, based on TCGA database and incorporated genomic and clinicopathologic factors, for the prediction of survival in patients with LSCC. The nomogram is significantly better than TNM stage alone in terms of the predictive value and clinical usability. Importantly, our nomogram presented good discrimination ability in early LSCC and advanced LSCC subgroups.

Abbreviations

LSCC: laryngeal squamous cell carcinoma; RNA-seq: RNA sequencing; TCGA: the Cancer Genome Atlas; lncRNAs: long non-coding RNAs; LASSO: least absolute shrinkage and selection operator; ROC: receiver operating characteristic; C-index: concordance index; OS: overall survival; AUC: area under curve; DCA: decision curve analysis; AJCC: American Joint Committee on Cancer; TNM: tumor-node-metastasis; LNR: lymph node ratio; LN: lymph nodes; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; IQRs: interquartile ranges; SDs: standard deviations; MI: multiple imputation; MCMC: Markov Chain Monte Carlo; MF: molecular function, BP: biological process; CC: cellular component; LNM: lymph node metastatic; GEO: Gene Expression Omnibus; DFS: disease-free survival; NSCLC: non-small cell lung cancer; OSCC: oral squamous cell carcinoma; HPV: human papillomavirus.

Declarations

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

All the data was obtained from TCGA, and the informed consent had been attained from the patients before our study.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are provided in supplementary materials and is also made available in the TCGA via the Genomics Data Commons <https://gdc.cancer.gov/>

Funding

Not applicable.

Authors' contributions

GLL, JC, and QQW conceived and designed the study. GLL, JSP, QW, YW, and LZH drafted the manuscript. GLL, JC, JSP, QZ and QQW analyzed and interpreted all the data. GLL, ZC, QZ, and JSP prepared the figures and tables. GLL, JC, QQW, and ZC reviewed and revised the manuscript. All authors have read and approved the manuscript for publication.

Authors' information

1 Department of Pathology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China.

2 Department of Head and Neck Surgery, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China.

3 Department of Intensive Care Unit, Shunde Hospital, Southern Medical University (The First people's hospital of Shunde), Foshan, 528308, Guangdong Province, PR China.

Acknowledgements

Not applicable.

References

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2018. 2018, 68(1):7-30.
2. Unsal AA, Kilic S, Dubal PM, Baredes S, Eloy JA: A population-based comparison of European and North American sinonasal cancer survival. *Auris, nasus, larynx* 2018, 45(4):815-824.
3. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F: Cancer incidence and mortality patterns in Europe: estimates for 4 countries in 2012. *European journal of cancer (Oxford, England : 1990)* 2013, 49(6):1374-1403.
4. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, Trama A, Visser O, Brenner H, Ardanaz E et al: Cancer survival in Europe 1999-2007 by country and age: results of EURO CARE-5-a population-based study. *The Lancet Oncology* 2014, 15(1):23-34.
5. Megwalu UC, Sikora AG: Survival outcomes in advanced laryngeal cancer. *JAMA otolaryngology-head & neck surgery* 2014, 140(9):855-860.
6. Lydiatt WM, Patel SG, O'Sullivan B, Brandwein MS, Ridge JA, Migliacci JC, Loomis AM, Shah JP: Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer stage manual. *CA: a cancer journal for clinicians* 2017, 67(2):122-137.
7. Jover-Espla AG, Palazon-Bru A, Folgado-de la Rosa DM, Severa-Ferrandiz G, Sancho-Mestre M, de Juan-Herrero J, Gil-Guillen VF: A predictive model for recurrence in patients with glottic cancer implemented in a mobile application for Android. *Oral oncology* 2018, 80:82-88.
8. Stein LD: Human genome: end of the beginning. *Nature* 2004, 431(7011):915-916.
9. Prensner JR, Chinnaiyan AM: The emergence of lncRNAs in cancer biology. *Cancer discovery* 2011, 1(5):391-407.
10. Tano K, Akimitsu N: Long non-coding RNAs in cancer progression. *Frontiers in genetics* 2012, 3:219.
11. Yarmishyn AA, Kurochkin IV: Long noncoding RNAs: a potential novel class of cancer biomarkers. *Frontiers in genetics* 2015, 6:145.
12. Liu G, Zheng J, Zhuang L, Lv Y, Zhu G, Pi L, Wang J, Chen C, Li Z, Liu J et al: A Prognostic 5-lncRNA Expression Signature for Head and Neck Squamous Cell Carcinoma. *Scientific reports* 2018, 8(1):15250.

13. Mao X, Qin X, Li L, Zhou J, Zhou M, Li X, Xu Y, Yuan L, Liu QN, Xing H: A 15-long non-coding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. *Gynecologic oncology* 2018, 149(1):181-187.
14. Zhu X, Tian X, Yu C, Shen C, Yan T, Hong J, Wang Z, Fang JY, Chen H: A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Molecular cancer* 2016, 15(1):60.
15. Tibshirani R: The lasso method for variable selection in the Cox model. *Statistics in medicine* 1997, 16(4):385-395.
16. Huang da W, Sherman BT, Lempicki RA: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 2009, 37(1):1-13.
17. Vickers AJ, Van Calster B, Steyerberg EW: Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ (Clinical research ed)* 2016, 352:i6.
18. Donders AR, van der Heijden GJ, Stijnen T, Moons KG: Review: a gentle introduction to imputation of missing values. *Journal of clinical epidemiology* 2006, 59(10):1087-1091.
19. Folch-Fortuny A, Villaverde AF, Ferrer A, Banga JR: Enabling network inference methods to handle missing data and outliers. *BMC bioinformatics* 2015, 16:283.
20. Shen Z, Li Q, Deng H, Lu D, Song H, Guo J: Long non-coding RNA profiling in laryngeal squamous cell carcinoma and its clinical significance: potential biomarkers for LSCC. *PloS one* 2014, 9(9):e108237.
21. Chen J, Shen Z, Deng H, Zhou W, Liao Q, Mu Y: Long non-coding RNA biomarker for human laryngeal squamous cell carcinoma prognosis. *Gene* 2018.
22. Zhao L, Cao H, Chi W, Meng W, Cui W, Guo W, Wang B: Expression profile analysis identifies the long non-coding RNA landscape and the potential carcinogenic functions of LINC00668 in laryngeal squamous cell carcinoma. *Gene* 2019, 687:47-55.
23. Feng L, Wang R, Lian M, Ma H, He N, Liu H, Wang H, Fang J: Integrated Analysis of Long Noncoding RNA and mRNA Expression Profile in Advanced Laryngeal Squamous Cell Carcinoma. *PloS one* 2016, 11(12):e0169232.
24. Bai Z, Shi E, Wang Q, Dong Z, Xu P: A potential panel of two-long non-coding RNA signature to predict recurrence of patients with laryngeal cancer. *Oncotarget* 2017, 8(41):69641-69650.
25. Yuan S, Xiang Y, Wang G, Zhou M, Meng G, Liu Q, Hu Z, Li C, Xie W, Wu N et al: Hypoxia-sensitive LINC01436 is regulated by E2F6 and acts as an oncogene by targeting miR-30a-3p in non-small cell lung cancer. *Molecular oncology* 2019, 13(4):840-856.
26. Yang L, Yi K, Wang H, Zhao Y, Xi M: Comprehensive analysis of lncRNAs microarray profile and mRNA-lncRNA co-expression in oncogenic HPV-positive cervical cancer cell lines. *Oncotarget* 2016, 7(31):49917-49929.
27. Li S, Chen X, Liu X, Yu Y, Pan H, Haak R, Schmidt J, Ziebolz D, Schmalz G: Complex integrated analysis of lncRNAs-miRNAs-mRNAs in oral squamous cell carcinoma. *Oral oncology* 2017, 73:1-9.
28. Miao Y, Sui J, Xu SY, Liang GY, Pu YP, Yin LH: Comprehensive analysis of a novel four-lncRNA signature as a prognostic biomarker for human gastric cancer. *Oncotarget* 2017, 8(43):75007-75024.

29. Saraniti C, Speciale R, Gallina S, Salvago P: Prognostic role of resection margin in open oncologic laryngeal surgery: survival analysis of a cohort of 139 patients affected by squamous cell carcinoma. *Brazilian journal of otorhinolaryngology* 2018.
30. Tassone P, Savard C, Topf MC, Keane W, Luginbuhl A, Curry J, Cognetti D: Association of Positive Initial Margins With Survival Among Patients With Squamous Cell Carcinoma Treated With Total Laryngectomy. *JAMA otolaryngology– head & neck surgery* 2018, 144(11):1030-1036.
31. Hou Y, Yu Z, Tam NL, Huang S, Sun C, Wang R, Zhang X, Wang Z, Ma Y, He X et al: Exosome-related lncRNAs as predictors of HCC patient survival: a prognostic model. *American journal of translational research* 2018, 10(6):1648-1662.
32. Egelmeier AG, Velazquez ER, de Jong JM, Oberije C, Geussens Y, Nuyts S, Kremer B, Rietveld D, Leemans CR, de Jong MC et al: Development and validation of a nomogram for prediction of survival and local control in laryngeal carcinoma patients treated with radiotherapy alone: a cohort study based on 994 patients. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 2011, 100(1):108-115.
33. Pan Y, Hong Y, Liang Z, Zhuang W: Survival analysis of distant metastasis of laryngeal carcinoma: analysis based on SEER database. *European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies* 2019, 276(1):193-201.
34. Imre K, Pinar E, Oncel S, Calli C, Tatar B: Predictors of extracapsular spread in lymph node metastasis. *European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies* 2008, 265(3):337-339.
35. Fletcher KT, Gal TJ, Ebelhar AJ, Valentino J, Brill YM, Dressler EV, Aouad RK: Prognostic indicators and survival in salvage surgery for laryngeal cancer. 2017, 39(10):2021-2026.
36. Chen AM, Felix C, Wang PC, Hsu S, Basehart V, Garst J, et al. Reduced-dose radiotherapy for human papillomavirus-associated squamous-cell carcinoma of the oropharynx: a single-arm, phase 2 study. *The Lancet Oncology*. 2017;18:803-11.
37. Ahmadi N, Ahmadi N, Chan MV, Huo YR, Sritharan N, Chin R: Laryngeal Squamous Cell Carcinoma Survival in the Context of Human Papillomavirus: A Systematic Review and Meta-analysis. *Cureus* 2018, 10(2):e2234.
38. Isa AY, Ward TH, West CM, Slevin NJ, Homer JJ: Hypoxia in head and neck cancer. *The British journal of radiology* 2006, 79(946):791-798.
39. Tuna M, Amos CI, Mills GB: Genome-Wide Analysis of Head and Neck Squamous Cell Carcinomas Reveals HPV, TP53, Smoking and Alcohol-Related Allele-Based Acquired Uniparental Disomy Genomic Alterations. *Neoplasia (New York, NY)* 2019, 21(2):197-205.
40. Te Riele R, Dronkers EAC, Wieringa MH, De Herdt MJ, Sewnaik A, Hardillo JA, Baatenburg de Jong RJ: Influence of anemia and BMI on prognosis of laryngeal squamous cell carcinoma: Development of an updated prognostic model. *Oral oncology* 2018, 78:25-30.

Tables

Table1. Characteristics of study population with number of missing values (n=109).

Variable	Category	No. (%) or median (IQR)	Missing values (%)
Age(years)		62(38-83)	0 (0)
Sex	Male	90(82.6)	0 (0)
Smoking history	Yes	46(42.2)	3(2.8)
Alcohol history	Yes	69(63.3)	2(1.8)
Number of lymph nodes		36(0-121)	18(16.5)
Number of positive LNs		1(0-42)	18(16.5)
Lymph node ratio		0.18(0-1)	18(16.5)
Margin status			15(13.8)
	Negative	84(77.1)	
	Positive	10(9.2)	
Tumor status			6(5.5)
	Tumor free	73(67)	
	Tumor with	30(27.5)	
Tumor grade			4(3.7)
	G1-G2	76(69.7)	
	G3-G4	29(26.6)	
Clinical T			4(3.7)
	T1-T2	19(17.4)	
	T3-T4	86(78.9)	
Clinical N			6(5.5)
	N0	53(48.6)	
	N1-N3	50(45.9)	
Clinical stage			4(3.7)
	I-II	13(11.9)	
	III-IV	92(84.4)	
Mutation count		153(36-861)	2(1.8)
Fraction genome altered		0.30(0-0.89)	1(0.9)

Abbreviations: IQR= interquartile range, LN= lymph node,

Table2. Univariable and multivariable Cox regression analysis for prediction of OS.

Factors	Subgroup	Univariable analysis		Multivariable analysis	
		HR(95%CI)	<i>P</i>	HR(95%CI)	<i>P</i>
Age		1.00 (0.97-1.04)	0.984	NA	NA
Sex	Female	1			
	Male	0.28(0.14-0.56)	0.000*	0.45(0.31-1.07)	0.090
Smoking history	No	1			
	Yes	0.65(0.35-1.18)	0.156	NA	NA
Alcohol history	No	1			
	Yes	0.77 (0.43-1.39)	0.388	NA	NA
Number of Lymph nodes		1.00 (0.98-1.01)	0.558	NA	NA
Number of positive LNs		1.00 (0.95-1.04)	0.892	NA	NA
Lymph node ratio		1.41 (0.28-7.13)	0.675	NA	NA
Margin status	Negative	1		1	
	Positive	4.68(2.08-10.51)	0.000*	3.00 (1.47-6.10)	0.003*
Tumor status	Tumor free	1		1	
	Tumor with	4.10(2.23-7.55)	0.000*	3.25 (1.79-5.90)	0.000*
Tumor grade	G1-G2	1			
	G3-G4	0.51(0.25-1.04)	0.064	NA	NA
Clinical T	T1-T2	1			

	T3-T4	0.72(0.35-1.50)	0.376	NA	NA
Clinical N	N0	1			
	N1-N3	1.44(0.80-2.57)	0.222	NA	NA
Clinical stage	I-II	1			
	III-IV	0.86(0.36-2.03)	0.729	NA	NA
Mutation count		0.99(0.98-1.01)	0.542	NA	NA
Fraction Genome altered		1.44(0.29-7.18)	0.654	NA	NA
LncRNAs signature		1.22(1.14-1.31)	0.000*	1.21(1.11-1.31)	0.000*

Abbreviations: HR=hazard ratio, CI =confidence intervals, OS=overall survival

NOTE: NA, not available. These variables were eliminated in the multivariate logistic regression model, so the HR and *P* values were not available. **P* < 0.05

Table3.Assessing the prognostic performance of the TNM stage, lncRNAs signature and nomogram

Model	Homogeneity monotonicity and discriminatory ability			
	Likelihood ratio(LR) test*	Linear trend χ^2 test**	C-index (95% CI) ***	Akaike information criterion(AIC)****
TNM stage	00.5	00.5	0.53 (0.45-0.61)	380.3
LncRNAs signature	21.4	46.9	0.78 (0.71-0.85)	355.0
Nomogram	48.0	71.7	0.82 (0.77-0.87)	332.4

*Higher homogeneity likelihood ratio indicates a smaller difference within the staging system, it means better homogeneity

**Higher discriminatory ability linear trend indicates a higher linear trend between staging system, it means better discriminatory ability and gradient monotonicity

***A higher c-index means better discriminatory ability.

****Smaller AIC values indicate better optimistic prognostic stratification

Figures

Figure 1

Thirteen lncRNAs selected by LASSO Cox regression analysis. (A) The two dotted vertical lines are drawn at the optimal values by minimum criteria (left) and 1 - s.e. criteria (right). Details are provided in Methods. (B) LASSO coefficient profiles of the 31 lncRNAs. A vertical line is drawn at the optimal value by 1 - s.e. criteria and results in thirteen non-zero coefficients. Thirteen lncRNAs—AC007907.1, AC025419.1, AC078993.1, AC090241.2, AL158166.1, AL355974.2, AL596330.1, HOXB-AS4, KLHL6-AS1, LHX1-DT, LINC00528, LINC01436 and TTTY14—with coefficients 0.2102, 0.0045, 0.1377, -0.3675, -0.0652, 0.0180, 0.1208, 0.0969, 0.2227, 0.1541, -0.0647, -0.0750, -0.1360., respectively, were selected in the LASSO Cox regression model.

Figure 2

Development of lncRNAs signature for prediction of survival in LSCC patients.(A and B) Distribution of lncRNAs-based classifier risk score. (C) Time-independent ROC curves with AUC values to evaluate predictive efficacy of lncRNA signature risk score. (D) Kaplan-Meier estimates of patients' survival status and time using the optimal lncRNA signature risk score cut-off which divided patients into low-risk and high-risk groups.

Figure 3

Functional annotation of the prognostic lncRNAs. Significantly enriched using the coexpressed mRNAs of the lncRNAs in GO analysis (A) and KEGG pathway analysis (B).

Figure 4

Nomogram for predicting 3-year and 5-year survival probability of LSCC after laryngectomy. To estimate risk, calculate points for each variable by drawing a straight line from patient's variable value to the axis labeled "Points." Sum all points and draw a straight line from the total point axis to the 3-year and 5-year survival axis.

Figure 5

ROC curves compare the prognostic accuracy of the nomogram with TNM stage or IncRNAs signature in predicting survival probability (A). Decision curve analysis for the nomogram, TNM stage and IncRNAs signature in prediction of prognosis of patients (B).

Figure 6

ROC curve analyses for survival prediction in subgroups of patients with LSCC. (A) advanced LSCC subgroup, and (B) early LSCC subgroup.

Figure 7

Kaplan-Meier estimates of patients' survival status and time using the optimal nomogram risk score cut-off which divided patients into low-risk and high-risk groups in subgroups of patients with LSCC. (A) advanced LSCC subgroup, and (B) early LSCC subgroup.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial2.xls](#)
- [Supplementarymaterial1.xls](#)