Oral Oncology 65 (2017) 94–101

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

A three-lncRNA signature derived from the Atlas of ncRNA in cancer (TANRIC) database predicts the survival of patients with head and neck squamous cell carcinoma

Wei Cao [a,b,1], Jian-nan Liu [a,b,1], Zeqi Liu [a,b,1], Xu Wang [a,b], Ze-Guang Han [c], Tong Ji [a,b,2,3], Wan-tao Chen [a,b,2,4], Xin Zou [c,*]

[a] Department of Oral Maxillofacial-Head and Neck Oncology, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China
[b] Shanghai Key Laboratory of Stomatology, Shanghai 200011, China
[c] Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Article history:
Received 6 September 2016
Received in revised form 12 November 2016
Accepted 17 December 2016

Keywords:
lncRNA
Signature
The Atlas of ncRNAs in Cancer
Predict
Survival
Head and neck squamous cell carcinoma

ABSTRACT

Objective: Long non-coding RNAs (lncRNAs) have important biological functions and can be used as prognostic biomarkers in cancer. To identify a lncRNA prognostic signature for head and neck squamous cell carcinoma (HNSCC).

Method: We analysed RNA-seq data derived from the TANRIC database to identify a lncRNA prognostic signature model using the orthogonal partial least squares discrimination analysis (OPLS-DA) and 1.5-fold expression change criterion methods. The prognosis prediction model based on the lncRNA signatures and clinical parameters were evaluated using the 5-fold cross validation method.

Results: A total of 84 out of 3199 lncRNAs were significantly associated with the survival of patients with HNSCC (log-rank test $P < 0.01$). Using the OPLS-DA and 1.5-fold change selection criterion, 5 lncRNAs (KTN1-AS1, LINC00460, GUSBP11, LINC00923 and RP5-894A10.6) were further selected. The prediction power of each combination of the 5 lncRNAs was evaluated through the receiver operating characteristic (ROC) curve and a three-lncRNA panel (KTN1-AS1, LINC00460 and RP5-894A10.6) achieved the highest prognostic prediction power (AUC 0.68, 95% CI 0.60–0.76, $P < 0.0001$) in the cohort. The patients were categorized into high- and low-risk groups based on their three-lncRNA profiles. Patients with high-risk scores had worse overall survival than those with low risk scores in the cohort (log-rank test $P = 0.0003$). Multivariable Cox regression analyses showed that the lncRNA signature and tumour grade were independent prognostic factors for patients with HNSCC.

Conclusions: Our findings showed that the three-lncRNA signature might be a novel biomarker for the accurate prognosis prediction of patients with HNSCC.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the most common malignant lesion in the head and neck region, with an estimated incidence rate of approximately 135.1 per 100,000 in China each year [1]. HNSCC is a heterogeneous solid tumour with more aggressive phenotypes and poor clinical outcomes primarily due to local tumour recurrence and regional lymph node and distant metastasis [2]. Therefore, better understanding of the genetic and epigenetic molecular alterations in HNSCC is the key to improving diagnosis, facilitating the development of appropriate treatments and promoting the prognosis of patients with HNSCC.

It has been demonstrated that approximately 70% of the human genome sequence can be actively transcribed, whereas more than

* Corresponding author at: B103, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China.
E-mail addresses: Jitong70@hotmail.com (T. Ji), chenwantao196323@sju.edu.cn (W.-t. Chen), x.zou@sjtu.edu.cn (X. Zou).
[1] The first three authors are equally contribute to this work.
[2] Co-Corresponding authors.
[3] Address: Department of Oral Maxillofacial-Head and Neck Oncology, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Floor 13, Building 1, 639 Zhizaoju Road, Shanghai 200011, China.
[4] Address: Department of Oral Maxillofacial-Head and Neck Oncology, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Floor 5, Building 8, 639 Zhizaoju Road, Shanghai 200011, China.

http://dx.doi.org/10.1016/j.oraloncology.2016.12.017
1368-8375/

98% of the human genome is part of the non-coding region [3,4]. A large proportion of transcripts are non-coding RNAs (ncRNA), including short ncRNAs and long ncRNAs. Long non-coding RNAs (lncRNAs) are transcripts of more than 200 nucleotides in length without protein-coding functions [5]. Accumulating evidence has revealed that lncRNAs act as key regulators by participating in gene regulation at the transcriptional, posttranscriptional and chromosomal levels [6] and are involved in large range of biological processes [7,8]. For example, antisense lncRNAs can hybridize to their specific pre-mRNA, which leads to alternatively spliced mRNAs or the production of endogenous siRNAs through Dicer [9]. LncRNAs also act as a sponge that can silence miRNA expression [10]. Moreover, lncRNAs interact with proteins, which results in dysregulated protein activity and cellular localization [8]. Additionally, lncRNAs can be transferred to small ncRNAs and act as endo-siRNAs or miRNAs [11]. Recent accumulating evidence also suggests that two micropeptides encoded by long noncoding RNAs regulate muscle function [12,13]. Similar to miRNAs, the aberrant alteration of lncRNA expression has also been detected in many human disease types, particularly in cancers [14–17]. Some lncRNAs have also been identified as predictors for survival in serious of cancer types [18–23]. However, the significance of the lncRNA signature for prognosis prediction in HNSCC patients remains unknown.

In this study, we integrated RNA-seq data from TCGA database and matched clinical information from a large cohort of HNSCC patients to develop a three-lncRNA panel prognostic risk prediction model. The reliability of the model was evaluated by using the 5-fold cross validation method.

Materials and methods

TCGA database and clinical information of patients with HNSCC

We acquired the lncRNA expression information from TANRIC (The Atlas of ncRNA in Cancer) (http://ibl.mdanderson.org/tanric/_design/basic/query.html), which originated from the TCGA HNSCC RNA-seq database (http://cancergenome.nih.gov/). The TANRIC data was released on 16/04/2015 (version 1.0.6). After retrieving the data, we annotated each of the samples according to their barcode ID based on the available clinical information, including overall survival (OS), age, gender, anatomic neoplasm subdivision, history of neoadjuvant treatment, history of other malignancy, lymphovascular invasion status, perineural invasion status, clinical stage, pathological lymph node status (pN), pathological tumour stage, extracapsular spread (ECS), smoking history, alcohol history, margin status, and tumour grade. The clinical characteristics of the HNSCC patients are summarized in Table 1. In this study, the patients with HNSCCs were divided into two groups (good survival and poor survival groups) based on the 3-year baseline survival status.

Table 1
Clinicopathological parameters of patients in the cohort.

Variables	No. of patients
Age	
⩾60	112
<60	80
Gender	
Male	135
Female	57
Anatomic neoplasm subdivision	
Oral cavity	123
Others	69
History of neoadjuvant treatment	
Yes	8
No	184
History of other malignancy	
Yes	9
No	183
Lymphovascular invasion status	
Yes	38
No	84
Unknown	70
Perineural invasion	
Yes	63
No	69
Unknown	60
Clinical stage	
I, II	46
III, IV	144
Unknown	2
pN	
Negative	54
Positive	129
Unknown	9
Pathological tumour stage	
I, II	40
III, IV	127
Unknown	25
ECS	
Negative	77
Positive	42
Unknown	73
Smoking history	
Lifelong non-smoker	37
Smoker	146
Unknown	9
Alcohol history	
No	68
Yes	120
Unknown	4
Margin status	
Negative	148
Positive	26
Unknown	18
Tumour grade	
G1	21
G2-4	170
Unknown	1

Abbreviations: ECS, extracapsular spread; pN, pathological lymph node status.

Identification of a potential prognostic lncRNA signature associated with OS and statistical analysis

The TANRIC resource contains lncRNA profiles of more than 10,000 features and 426 HNSCCs. Due to the very low expression level of lncRNAs in HNSCCs, ultimately only 3199 lncRNAs in HNSCCs were acquired from the TANRIC resources in this study. The log-rank test and univariate Cox regression analysis were first carried out to evaluate the association between the lncRNAs expression levels and the OS of the patients with HNSCC. At a confidence level of $P < 0.01$, 84 out of 3199 lncRNAs were identified to be significantly associated with survival in HNSCC patients (Supplementary Table 1). In this study, we defined "good survival" as being survival no less than three years and the patients with a survival time less than three years were categorized as "poor survival". The patients with a survival status of no more than three years or the patients who died after more than three years were excluded. Therefore, we finally incorporated 192 out of 426 patients in the subsequent analyses. In the lncRNA signature analyses, the good survival group included 75 patients and the poor survival group contained 117 patients. The OPLS-DA analysis was