Original Article

Construction of an eight-gene signature for survival evaluation of papillary thyroid cancer patients

Haiguang Liu¹, Yehuan Liu²*, Jinmiao Qu¹

Departments of ¹Thyroid & Breast Surgery, ²General Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, People’s Republic of China. *Equal contributors.

Received December 17, 2018; Accepted April 8, 2019; Epub June 15, 2019; Published June 30, 2019

Abstract: Background: Papillary thyroid carcinoma (PTC), accounting for 80% of all thyroid cancer cases, is the most frequent type of thyroid cancer. It is critical in identifying novel disease biomarkers, thus improving PTC patient management and making tailored therapeutic decisions. Methods: RNA-sequencing and survival data of PTC patients and normal thyroid tissues were retrieved from the Cancer Genome Atlas (TCGA). Univariate Cox analysis was utilized to investigate the relationship between expression levels of each differentially-expressed gene and survival. Multivariate Cox analysis was utilized for genes significantly associated with overall survival. With the co-efficient values of multivariate analysis, a gene prognostic signature was constructed by calculating risk scores for each subject. Results: A total of 1,757 genes were selected as significant differentially-expressed genes in PTC, compared to normal thyroid tissues. Fifteen genes were shown to be significantly associated with overall survival of PTC patients, according to univariate analysis. Multivariate Cox proportional hazards regression analysis recognized 8 differentially-expressed genes remarkably correlated with overall survival. Weighted by the corresponding co-efficient, an eight-gene prognostic signature was constructed: risk score = (0.3930)*Exp METTL7B + (-0.5943)*Exp ADRA1B + (-0.6605)*Exp RIPPLY3 + (-0.5314)*Exp FAM111B + (0.5030)*Exp PCOLCE2 + (-0.7236)*Exp LINC01208 + (-0.5016)*Exp ZSCAN4 + (0.4953)*Exp SALL3. The eight-gene signature exhibited strong predictive power for 5-year survival of PTC patients. The area under the time-dependent ROC curve was 0.969. Conclusion: An eight-gene signature was identified. It can be used as an independent prognostic marker, robustly predicting survival of PTC patients.

Keywords: Papillary thyroid cancer, survival, gene signature

Introduction

Thyroid cancer is the most common endocrine malignancy [1, 2]. In the past few decades, incidence of thyroid cancer has increased by 4%, worldwide, becoming the fastest growing type of cancer in many countries [3, 4]. It was estimated that there were 53,990 new cases of thyroid cancer in the United States in 2018, with 2,060 deaths [5]. Generally, thyroid cancer may be divided into four pathological types, including papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), anaplastic thyroid carcinoma (ATC), and medullary thyroid carcinoma (MTC), according to histopathological examinations. PTC, accounting for 80% of all thyroid cancer cases, is the most frequent type of thyroid cancer [6]. Overall, treatment with surgery, thyroid hormone therapy, and radioiodine therapy is effective. Most PTC patients have a good prognosis, with a 10-year survival rate close to 90% [7]. However, some PTC patients suffer from cancer recurrence. Some PTCs may develop distant metastases with high mortality rates. According to previous studies, local recurrence occurs in up to 20% of patients with thyroid cancer. Nearly 10% of patients with thyroid cancer will have distant metastases within 10 years [8]. Hence, it is critical to identify novel disease biomarkers, thus improving PTC patient management and making tailored therapeutic decisions.

In recent years, molecular biology has greatly improved concerning prognosis of PTC. Some molecular biomarkers have been proposed for risk stratification of PTC, as well as traditional clinical pathology parameters. Typical examples
are BRAF and telomerase reverse transcriptase (TERT). The coexistence of BRAF V600E and TERT promoter mutations indicates aggressive phenotypes [9, 10]. A deeper understanding of the drivers and molecular mechanisms of PTC tumorigenesis, as well as the recording of candidate prognostic markers, is critical for better PTC diagnostic and therapeutic strategies.

In the current study, transcriptome data of PTC was obtained from the Cancer Genome Atlas (TCGA). Expression levels of genes were quantified. Next, this study scanned novel gene biomarkers closely associated with survival of PTC patients, building a gene signature for prediction of survival in patients with PTC.

Materials and methods

Patient cohort and identification of differentially-expressed genes

RNA-sequencing and survival data of 510 PTC patients and 58 normal thyroid tissues were retrieved from TCGA (http://cancergenome.nih.gov/). Differentially-expressed genes were examined using the edgeR package in R software. Differentially-expressed genes between PTC tissues and normal thyroid tissues were selected using the following criteria: 1) Fold change (FC) > 4 for upregulation or downregulation; and
2) False discovery rate (FDR) < 0.05. The heatmap and volcano plot were drawn using differentially-expressed gene analysis. Color was determined by filtering criteria. Clinical information concerning these 510 patients was downloaded from TCGA database. Significant differentially-expressed genes, as well as clinical characteristics of patients, are shown in Tables S1, S2.

Construction of the gene prognostic signature

Univariate Cox analysis was used to study the relationship between expression levels and survival of each of the differentially-expressed genes. Genes with \( P \) values < 0.001 were selected as candidate gene biomarkers. Subsequently, to ensure that each candidate gene biomarker was used as an independent indicator of survival, multivariate Cox proportional hazards regression analysis was used for genes significantly associated with overall survival, according to univariate analysis. A gene prognostic signature was constructed by calculating risk scores for each subject using co-efficient values obtained through multivariate analysis. To identify the distinguishing ability of patient outcomes, Kaplan-Meier survival curves were conducted. Time-dependent receiver operating characteristic (survival ROC) curves were also employed to detect the prognostic power of the risk score model.

Statistical analysis

Clinicopathological characteristics were evaluated using Chi-squared tests. Kaplan-Meier and Cox regression analyses were utilized to assess the association between the risk score model and overall survival of PTC. Statistical analyses were performed using SPSS version 20.0 (Chicago, IL).

Results

Selection of differentially-expressed genes

Initially, detailed information concerning 510 PTC patients and 58 normal thyroid tissues was obtained from the TCGA dataset. Compared with normal thyroid tissues, 1,757 genes were selected as differentially-expressed genes in PTC. Of these, 402 genes were downregulated and 1,355 genes were upregulated. An overview of aberrantly-expressed genes is shown in Figure 1. All differentially-expressed genes are listed in Table S3.

Specific differentially-expressed genes correlated with overall survival

Regarding the correlation between differentially-expressed genes and overall survival, 15 differentially-expressed genes were evaluated, with univariate analysis, and found to be significantly associated with overall survival of PTC patients (Table 1). These 15 differentially-expressed genes were subjected to further analysis via multivariate analysis. Multivariate Cox proportional hazards regression analysis recognized 8 differentially-expressed genes that were remarkably correlated with overall survival (OS, Table 2). Weighted by corresponding co-efficient values, an eight-gene prognostic signature was constructed: risk score = \((0.3930)\times\text{Exp METTL7B} + (-0.5943)\times\text{Exp ADRA1B} + (-0.6605)\times\text{Exp RIPPLY3} + (-0.5314)\times\text{Exp FAM111B} + (0.5030)\times\text{Exp PCOLCE2} + (-0.7236)\times\text{Exp LINC01208} + (-0.5016)\times\text{Exp ZSCAN4} + (0.4953)\times\text{Exp SALL3}\).

As shown in the risk score model, co-efficient values of ADRA1B, RIPPLY3, FAM111B, LINC01208, and ZSCAN4 were negative, indicating a positive correlation with OS of PTC patients. Expression of METTL7B, PCOLCE2, and SALL3 was shown to be a negative factor.

Evaluating the predictive power of the risk model, the patients were divided into two groups (high- and low-risk score groups) using median risk score values as the cutoff. The supervised heatmap showed expression levels of these eight-genes between the low-risk score model.
group and high-risk score group (Figure 2). As shown in Figure 3, differences between the survival curves of the two groups were statistically significant (P < 0.001). Cumulative 3, 5, and 10-year OS rates in the low-risk score group were all maintained at 100%, while OS rates were 94.262%, 84.317%, and 79.357%, respectively, in the high-risk score group. Furthermore, the eight-gene signature exhibited strong predictive power for 5-year survival of PTC patients (the area under the ROC curve was 0.969, Figure 4). Results suggest that the risk score model provided good sensitivity and specificity in predicting PTC patient survival.

Correlation between clinical characteristics and the risk model

To examine the association between the eight-gene signature and clinical characteristics of PTC patients, further analysis was performed. Patients were categorized as the low-risk group or high-risk group depending on median risk score values. Results revealed that age (P = 0.001), N stage (P = 0.003), and histological type (P = 0.023) were significant related with the eight-gene signature risk model (Table 3). PTC patients with a high risk were inclined to be older and have the follicular type.

Discussion

Thyroid cancer is a common malignant endocrine disorder, worldwide, with increasing incidence rates. According to worldwide cancer reports, there were 567,233 new thyroid cancer cases in 2018, with about 41,071 patients dying [11]. PTC, the most frequent type of thyroid cancer, has a relatively slow progression and favorable prognosis. However, approximately 5-20% of PTC patients still face the challenge of cancer recurrence, suffering fatal outcomes. Therefore, identifying new disease biomarkers may help to improve PTC patient management and develop customized treatment.
decisions. The current study analyzed transcriptome data of PTC in a large-scale cohort, showing altered expression patterns between tumor and non-tumor tissues. More than 1,700 genes, which contain mRNA and IncRNA, were identified as differentially-expressed genes. On this basis, univariate Cox proportional hazard regression and multivariate Cox proportional hazard analysis models were used to identify genes with expression levels that significantly correlated with PTC patient survival. Finally, a signature comprised of eight genes (METTL7B, ADRA1B, RIPPLY3, FAM111B, PCOLCE2, LINC01208, ZSCAN4, and SALL3) was identified. It can be used as an independent prognostic marker to robustly predict survival rates of patients with PTC.

Genetic diagnosis has been a great success in the field of cancer. Typical examples are 21-Gene and 70-Gene signatures in breast cancer, providing clinically useful prognostic information and treatment options [12-15]. Previous studies have developed several diagnostic signatures in PTC. Brennan et al. discovered a gene expression signature, containing 109 genes, that distinguishes extremely good from extremely poor prognosis in patients [16]. As a classification to predict prognosis using extremely poor and good prognosis patients, the AUC value of the gene expression signature was 0.75. Xin You et al. developed a three-IncRNA signature (PRSS3P2, KRTAP5-AS1, and PWAR5) [17]. Their study showed that PTC patients with low-risk scores tended to gain obviously longer survival times. The area under the time-dependent ROC curve was 0.739. Moreover, another group reported that four IncRNAs, including RP11-536-N17.1, RP11-508M8.1, AC02-6150.8, and CTD-2139B15.2, were markedly related to progression and survival of PTC. This four-IncRNA signature was an independent biomarker predicting recurrence of PTC patients, with an AUC value of 0.833 [18]. However, these signatures should be tested concerning sensitivity and specificity values in a larger sample validation cohort before being used in clinical decision making.

All eight genes (METTL7B, ADRA1B, RIPPLY3, FAM111B, PCOLCE2, LINC01208, ZSCAN4, and SALL3) were significantly differentially-expressed between thyroid cancer tissues and non-cancerous thyroid tissues. This suggests that those eight genes may be potential oncogenes or anti-oncogenes in PTC. SALL3 gene encodes a sal-like C2H2-type zinc-finger protein and belongs to a family of evolutionarily conserved genes. SALL3 protein plays a role in embryonic development and DNA methylation [19-21]. A previous study showed that the SALL3 gene has a role in tumorigenesis of head and neck squamous cell carcinoma (HNSCC). Epigenetic silencing of SALL3 was shown to be an independent predictor of poor survival in HNSCC [22]. Another study reported that SALL3 was associative with aberrant methylation in human hepatocellular carcinoma [23]. ADRA1 is a member of the G protein-coupled receptor superfamily. It activates mitogenic responses and regulates growth and proliferation of many
cells. It has been reported that ADRA1 was associated with gastric cancer and breast cancer [24, 25]. ETTL7B gene encodes a methyltransferase like protein. It has been reported that METTL7B protein resides on cell membranes and is associated with sumoylation regulation [26, 27]. To the best of our knowledge, no other groups have reported the biological function of these eight genes. Further investigation into the function of these genes may provide additional targeting strategies for treatment.

The current study had several limitations, however. First, results should be validated with a larger sample size. With further independent validation studies in prospective cohorts, the current eight-gene prognostic signature might serve as an alternative biomarker and therapeutic target for PTC patients. Second, the functions of these eight genes have yet to be determined. Further in vitro and in vivo experiments are currently being conducted by the present group, investigating the biological effects of the abovementioned genes.

In conclusion, the current study performed genome-wide analysis of genomic expression in a large cohort of PTC patients from TCGA. Results showed changes in expression patterns between tumor and non-tumor samples. An eight-gene signature was identified. It can be used an independent prognostic marker, robustly predicting survival rates of PTC patients.

Acknowledgements

This study was funded by the Natural Science Foundation of Zhejiang Province (Number LY-17H160053, LGF18H160032, LGF18H160031, and LY18H16-0053).

Disclosure of conflict of interest

None.

Address correspondence to: Jinmiao Qu, Department of Thyroid & Breast Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang, People’s Republic of China. Tel: +86 577 5557 9462; E-mail: qujinmiao0577@163.com

References


Signature for PTC patients


