Peer Review File

Article information: https://dx.doi.org/10.21037/gs-21-361

Reviewer A

The paper does present the information in a straightforward and understandable manner. The scientific methods and assumptions are valid and clearly outlined. The description of data and the calculations are sufficiently complete to be followed and would allow their reproduction by fellow scientists. This is the strength of the paper. This reviewer has a few minor comments that are listed below.

Comment 1: Baseline demographic and clinical characteristics for all and each dataset should be reported in a table.

Reply 1: Thank you very much for your constructive comments which could enrich our manuscript. The baseline demographic and clinical characteristics for all, training and validation datasets were presented in a separate paragraph in **Results** section and a supplementary table (**Table 1**) in our revised manuscript.

Results, Paragraph 1: Clinical Characteristics of PTC patients in TCGA Database

In total, there were 443 samples with data on the N stage, comprising 226 samples in stage N0 (51.01%), 87 samples in stage N1a (19.64%),73 samples in stage N1b (16.48%), and 57 samples without further stratification as N1a or N1b (12.87%). According to the lymph node status, the 443 samples were divided into a training set (N=311) and an internal validation set (N=132). The baseline clinical characteristics are presented in Table 1. There was no significant difference in the status of LNM between the training and validation sets.

Table 1 Clinical Characteristics of PTC patients in TCGA database [n (%)]

Variables	Total (N=443)	Training Set (N=311)	Validation Set (N=132)	Statistics	*p* Value
Age*	47 (35,58)	47 (36,58)	46 (33,58)	-0.694	0.487
Sex				0.131	0.718
Female	324 (73.1)	229 (73.6)	95 (72.0)		
Male	119 (26.9)	82 (26.4)	37 (28.0)		
Number of Lymph Node Examined*	5 (2,16)	5 (2,15)	7 (3,19.5)	-1.142	0.254
T Stage				2.609	0.625
T1	131 (29.6)	98 (31.5)	33 (25.0)		
T2	139 (31.4)	94 (30.2)	45 (34.1)		
T3	150 (33.8)	102 (32.8)	48 (36.4)		
T4	22 (5.0)	16 (5.2)	6 (4.5)		
TX	1 (0.2)	1 (0.3)	0		

N Stage				1.261	0.738
N0	226 (51.0)	159 (51.1)	67 (50.8)		
N1	57 (12.9)	37 (11.9)	20 (15.1)		
N1a	87 (19.6)	61 (19.6)	26 (19.7)		
N1b	73 (16.5)	54 (17.4)	19 (14.4)		
Multifocality				0.125	0.940
Unifocal	233 (52.6)	165 (53.1)	68 (51.5)		
Multifocal	201 (45.4)	140 (45.0)	61 (46.2)		
Unknown	9 (2.0)	6 (1.9)	3 (2.3)		
Tumor Side				7.069	0.029
Unilateral	357 (80.6)	249 (80.1)	108 (81.8)		
Bilateral	81 (18.3)	61 (19.6)	20 (15.2)		
Unknown	5 (1.1)	1 (0.3)	4 (3.0)		
Radiation Therapy				3.449	0.178
No	160 (36.1)	115 (37.0)	45 (34.1)		
Yes	266 (60.1)	181 (58.2)	85 (64.4)		
Unknown	17 (3.8)	15 (4.8)	2 (1.5)		

*Age and number of lymph node examined are abnormally distributed continuous variables and represented by the median and upper and lower quartiles. The statistical significance was estimated by Mann-Whitney U test.

Comment 2: Please specify the edition of the AJCC staging system used in this study.
Reply 2: This is a very important and valuable suggestion which really make our manuscript more accurate. age at initial pathologic diagnosis, sex, number of lymph nodes examined, primary neoplasm focus type, primary thyroid gland neoplasm location anatomic site, pathologic TNM stage (which is defined following the AJCC 7th edition), radiation therapy status, and disease-free survival (DFS). We have added these contents in **Methods** section in our revised manuscript.

Methods, paragraph 1: Datasets

The transcriptome data (FPKM) of 568 thyroid carcinoma samples in The Cancer Genome Atlas (TCGA) database were collected. The patients' clinical characteristics and survival data of all the PTC samples were obtained from UCSC Xena (https://xena.ucsc.edu; University of California, Santa Cruz). Clinical characteristics data including age at initial pathologic diagnosis, sex, number of lymph nodes examined, primary neoplasm focus type, primary thyroid gland neoplasm location anatomic site, pathologic TNM stage (which is defined following the AJCC 7th edition), radiation therapy status, and disease-free survival (DFS). The clinical data were re-evaluated according to the original pathologic reports. Cases of unknown lymph node status and non-PTC samples were excluded. Samples were divided into a training set (70%) and an internal validation set (30%) according to the status of cervical LNM. R

software (version 4.0.3) was used for data collection and processing.

Comment 3: There is no statement about research ethics.
Reply 3: We totally understand your concern. Our present study was based upon open-source data obtained from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga), which belongs to a public database. The patients involved in the database have given ethical approval. Users can download relevant data for free for research and publish relevant articles. So, we consider that there might be no ethical issues or other conflicts of interest. We have added a detailed statement in **Footnote/ Ethical Statement** section. We hope theses can resolve your concerns.
Footnote
Ethical Statement
Our present study was based upon open-source data obtained from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga), which belongs to a public database. The patients involved in the database have given ethical approval. Users can download relevant data for free for research and publish relevant articles. There are no ethical issues or other conflicts of interest. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Comment 4: The English quality and grammar of the text is uneven, the manuscript should benefit from language editing
Reply 4: We totally understand your concern. The revised manuscript has been checked by AME Editing Service. We hope the revised version will meet the Journal needs.

Reviewer B
Comment 1: The manuscript analyzed the TCGA database of PTC and found that 14 novel signatures are associated with lymph node metastasis and DFS. The study design is clear and logical. From my part, I have no further questions, and I think the manuscript could be published after language modification.
Reply 1: We totally understand your concern. The revised manuscript has been checked by AME Editing Service. We hope the revised version will meet the Journal needs.