



Identification and validation of a glycolysis-related gene signature for predicting metastasis and survival rate in patients with thyroid cancer

Bo Wang^{1#}, Yongqian Zhu^{2#}, Xiang Zhang^{3#}, Zijie Wang⁴

¹The Second Clinical Medical School, Nanjing Medical University, Nanjing, China; ²Department of Medical Quality Management, the First Affiliated Hospital with Nanjing Medical University, Nanjing, China; ³Department of Urology, Affiliated Hospital of Nantong University, Nantong, China; ⁴The First Clinical Medical School, Nanjing Medical University, Nanjing, China

Contributions: (I) Conception and design: X Zhang, B Wang; (II) Administrative support: Y Zhu; (III) Provision of study materials or patients: Y Zhu; (IV) Collection and assembly of data: B Wang, Z Wang; (V) Data analysis and interpretation: B Wang, Z Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Zijie Wang, PhD. The First Clinical Medical School, Nanjing Medical University, 300# Guangzhou Road, Nanjing, China. Email: njmuwzj@126.com.

Background: Thyroid cancer (THCA) is one of the most commonly malignant endocrine tumors worldwide. This study aimed to explore new gene signatures to better predict the metastasis and survival rate of patients with THCA.

Methods: mRNA transcriptome data and clinical characteristics of THCA were obtained from the Cancer Genome Atlas (TCGA) database to identify the expression and prognostic implications of glycolysis-related genes. Gene Set Enrichment Analysis (GSEA) of differentiated expressed genes was performed, and relationship between genes and glycolysis was observed in the Cox proportional regression model. Using the cBioPortal, mutations were subsequently identified in model genes.

Results: A 3-gene (*HSPA5*, *KIF20A* and *SDC2*) signature based on glycolysis related genes was identified and used to predict metastasis and survival rate in patients with THCA. Further expression analysis revealed that *KIF20A* was a poor prognostic gene, while *HSPA5* and *SDC2* were good prognostic genes. Evaluating the prognosis of patients with THCA could be more effective with this model.

Conclusions: The study reported a three-gene signature of THCA, including *HSPA5*, *KIF20A* and *SDC2*, which were found to be closely correlated with the glycolysis of THCA, and it showed a high efficacy to the prediction of metastasis and survival rate of THCA.

Keywords: Thyroid cancer (THCA); glycolysis; prognosis; metastasis; survival rate

Submitted Nov 04, 2022. Accepted for publication Apr 19, 2023. Published online May 12, 2023.

doi: 10.21037/tcr-22-2548

View this article at: <https://dx.doi.org/10.21037/tcr-22-2548>

Introduction

Thyroid cancer (THCA) is one of the most commonly malignant endocrine tumors worldwide. THCA currently ranks as the 13th most common cancer diagnosis overall and the 6th most common among women (1). In 1990, an estimated 95,030 cases of THCA were diagnosed and 22,070 deaths, were reported, and the number of incidents

increased to 255,490 incident cases and 41,240 deaths in 2017 (2). In 2020, THCA ranked 9th in incidence of all cancer types, with more than 586,000 cases diagnosed worldwide. The incidence of THCA in China is also on the rise and has become one of top ten cancers threatening the health of the public (3).

THCA is divided into four categories, among which

thyroid papillary carcinoma (PTC) accounts for more than 85% (4,5). Although an overwhelming majority of PTCs can be cured by surgical treatments and I¹³¹ treatments, 10% of the patients still die of poorly differentiated and advanced tumors (6). To a large extent, the increasing number of THCAs is attributed to the advancement of the progressively sensitive diagnostic imaging modalities (7,8). Therefore, the compelling need for early and accurate diagnosis and more effective treatment strategies are accompanied.

It is believed that the development and prognosis of THCA are related to the mutation and expression of some genes. For instance, lipid metabolism-related genes play an important role in the prognosis of THCA, and the risk of THCA has also been linked to impaired glucose metabolism (9,10). However, the pathogenesis of THCA is complicated and ambiguous, which requires further study.

In recent years, the metabolic process of tumor environment has become crucial in tumor research and treatment (11,12). Furthermore, the discovery of Warburg effect, which was introduced by German scientist Otto Warburg, is a vital breakthrough in cancer research (13,14). Cancer cells rewire their metabolism to promote growth, survival, proliferation, and long-term maintenance. The common feature of this altered metabolism is the increased glucose uptake and fermentation of glucose to lactate. This phenomenon is observed even in the presence of completely functioning mitochondria and, together, is known as the 'Warburg Effect'. Most of the hallmarks of cancer could be the consequence of the Warburg's effect (15,16). Therefore,

the relationship between gene signatures and THCA was explored in this study by identifying glycolysis-related genes (GRGs) according to Warburg effect.

The purpose of this study was to identify the GRGs in the progression and metastasis of THCA. Gene Set Enrichment Analysis (GSEA) was used to select differentiated expressed genes. Finally, a 3-GRG risk signature was established which could effectively predict patient prognosis. Moreover, the gene-based model in the study, as an independent prediction factor, could identify patients with a high-risk score who showed a poorer prognosis than those with a low-risk score. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-22-2548/rc>).

Methods

Data acquisition

mRNA transcriptome data of THCA were obtained from TCGA. The RNA sequence data of THCA were downloaded directly from TCGA data portal (<https://portal.gdc.cancer.gov/>). The disease types are all adenomas and adenocarcinomas. Data of Glycolysis-associated gene sets were obtained from GSEA official website. Data showing missing information or being inconsistent were excluded from the analysis. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

GSEA analysis of glycolysis-related pathways

Data of Glycolysis-associated gene sets were obtained from the official website of the GSEA. The pathways were: GO_GLYCOLYTIC_PROCESS (GGP), KEGG_GLYCOLYSIS_GLUONEOGENESIS, REACTOME_GLYCOLYSIS, BIOCARTEA_GLYCOLYSIS_PATHWAY, and HALLMARK_GLYCOLYSIS (HG). For GSEA analysis, the parameter for gene sets permutations was set 1,000 times, and a false discovery rate (FDR) <0.05 was considered significant.

Differential Expression Analysis and Glycolysis Related Genes Model

Based on the results of GSEA, glycolytic pathway with a close relationship with THCA (P<0.05) was screened. In these pathways, GRGs were summarized using the expression levels of genes. Then, using the limma software package in R x64 v3.6.3 software, the expression level of

Highlight box

Key findings

- A 3-gene (HSPA5, KIF20A and SDC2) signature was identified and used to predict metastasis and survival rate in patients with thyroid cancer (THCA).

What is known and what is new?

- Glycolysis is known to play an important role in the development and metastasis of THCA.
- The study reported a three-gene signature of THCA, which were found to be closely correlated with the glycolysis of THCA, and it showed a high efficacy to the prediction of metastasis and survival rate of THCA.

What is the implication, and what should change now?

- This study highlights the importance of glycolysis-related genes in clinical practice and provides a new method for the assessment of THCA patients.

Table 1 TCGA clinical information

Variable	Count
Age (years)	
≤65	436
>65	71
Gender	
Female	371
Male	136
Stage	
I-II	337
III-IV	168
Missing	2
Stage T	
T1-T2	311
T3-T4	194
Missing	2
Stage N	
N0	231
N1-3	226
Missing	50
Stage M	
M0	283
M1	9
MX	215

TCGA, The Cancer Genome Atlas.

GRGs was extracted. Between glycolysis pathway and the extracted GRGs, the dramatic differences ($P < 0.05$, $\log_{2}FC \geq 1$ or ≤ -1) in the expression profiles were analyzed. Cox risk regression analysis was used to further analyze the close correlation between GRGs and the survival rate, and the most significant genes were selected to construct the GRGs model. Receiver operating characteristic (ROC) curve was used to evaluate the accuracy and specificity of the model and the optimal cutoff point was obtained. Finally, as a basis for distinguishing the high-risk and the low-risk groups, the model's critical value was used.

Mutation of model genes

A Human Protein Atlas (HPA) (www.proteinatlas.org) search

was conducted for immunohistochemical images of the three genes in normal thyroid tissues and THCA tissues, and, by immunohistochemical staining, differences in gene expression between the two groups were confirmed. Mutation function in the cBioPortal website (www.cbioportal.org) was used to explore mutations in the 3 model genes.

Verification of GRGs model

The GRGs model was verified by univariate and multivariate COX proportional hazard analysis and survival analysis of THCA samples obtained from TCGA. Leave-one-out cross validation was performed to evaluate the accuracy of the model.

Statistical analysis

Data were presented as mean \pm standard deviation (SD) except when being stated otherwise, and compared using the student's *t*-test. Kaplan-Meier method was used to perform subsequent survival analysis and plot the graph survival curve. Univariate and multivariate COX regression analyses were used to filter differential expressed genes (DEGs) that significantly correlated with the prognosis of THCA (P value filter < 0.05). P values less than 0.05 were considered statistically significant. Risk score was calculated with help of the 'predict ()' function in the survival package in the R software version 3.6.3 (www.r-project.org).

Results

GSEA analysis of glycolysis-related pathways

The obtained expression data of 12,398 genes from 567 samples including 58 normal tissues and 509 tumor tissues were used for follow-up research. At the same time, 507 patients were classified. Among these patients, 71 were over 65 years old and 436 were under 65 years old; 371 women and 136 men. The clinical demographics of patients from this cohort are presented in *Table 1*. GSEA analysis revealed that the genes were significantly enriched ($P < 0.05$) in the GGP (*Figure 1A*) and the HG (*Figure 1B*) (*Table 2*).

Differential expression analysis and glycolysis related genes model

The extract of genes enriched in iconic glycolysis pathways

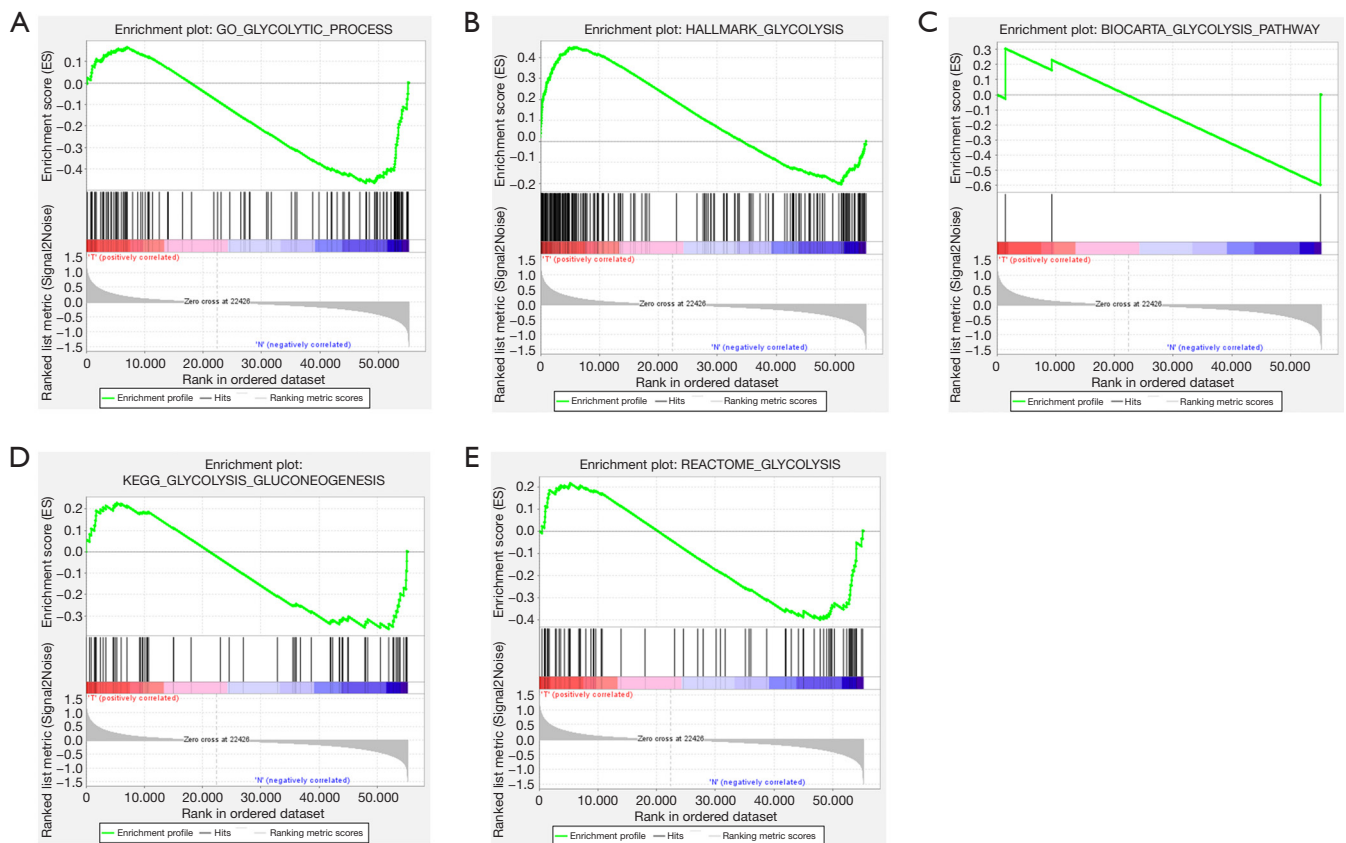


Figure 1 Results of GSEA analysis. (A) GO_GLYCOLYTIC_PROCESS. (B) HALLMARK_GLYCOLYSIS. (C) BIOCARTA_GLYCOLYSIS_PATHWAY. (D) KEGG_GLYCOLYSIS_GLUconeogenesis. (E) REACTOME_GLYCOLYSIS. GSEA, Gene Set Enrichment Analysis.

Table 2 Details of GSEA results

Name	Size	ES	NES	NOM P value	FDR q value
GO_GLYCOLYTIC_PROCESS	106	-0.46	-1.60	0.025	0.025
HALLMARK_GLYCOLYSIS	200	0.45	1.57	0.027	0.027
BIOCARTA_GLYCOLYSIS_PATHWAY	3	-0.60	-0.96	0.596	0.596
KEGG_GLYCOLYSIS_GLUconeogenesis	62	-0.36	-1.16	0.286	0.286
REACTOME_GLYCOLYSIS	72	-0.40	-1.30	0.178	0.178

GSEA, Gene Set Enrichment Analysis; ES, enrichment score; NES, normalized enrichment score; NOM, nominal; FDR, false discovery rate.

and glycolytic process was conducted for differential expression analysis. A total of 52 differentially expressed GRGs were identified (Table S1). Three genes (*HSPA5*, *KIF20A*, *SDC2*) closely related to clinical prognosis were identified by COX risk regression analysis, and they are described in detail in Table 3. Then, the prognostic model

was constructed. This prognostic model including three genes. It was found that *HSPA5* and *SDC2* were good prognostic genes in this model, whereas *KIF20A* was a bad prognostic gene. The area under curve (AUC) value was 0.68 (Figure 2A), based on the ROC curve of the model, indicating the moderate accuracy of the model. In the

Table 3 Three genes related to clinical prognosis identified by COX risk regression analysis

Name	Coef.	Hazard ratio	Physiological role
<i>HSPA5</i>	-0.00243	0.997568	The protein encoded by HSPA5 is a member of the HSP70 family; it is localized in the lumen of the ER, and is involved in the folding and assembly of proteins in the ER
<i>KIF20A</i>	0.641416	1.899168	KIF20A, a kinesin with unique structural features, has been proved to be potentially associated with prostate and liver cancer
<i>SDC2</i>	-0.00799	0.992038	SDC2 methylation is a feasible biomarker for colorectal cancer detection

Coef., coefficient; HSP70, heat shock protein 70; ER, endoplasmic reticulum.

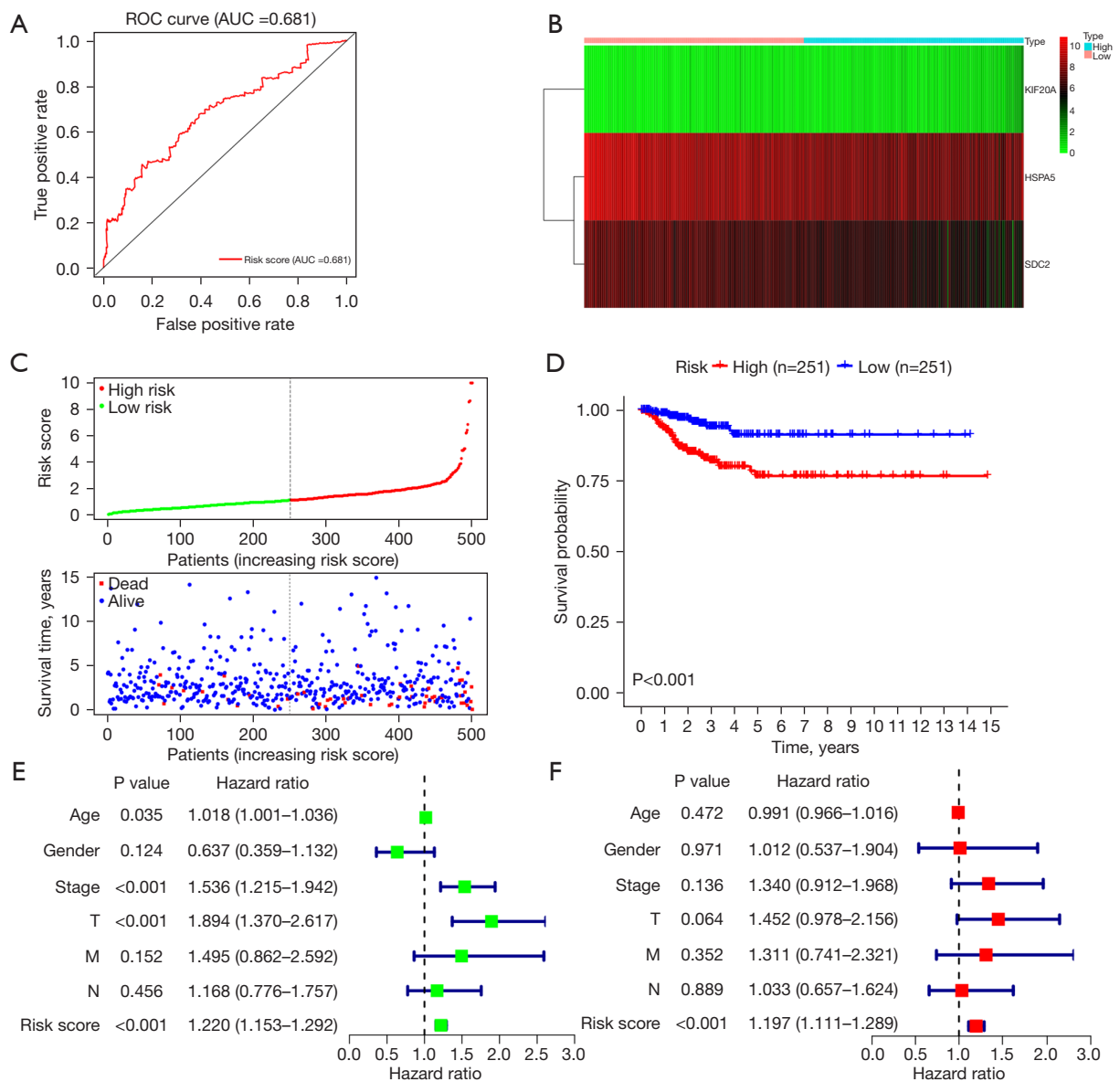


Figure 2 Construction and verification of prognostic model identified by high and low risk groups. (A) Receiver operating characteristic curve. (B) Relative expression of model gene. (C) The model divides patients into low-risk or high-risk groups. (D) Kaplan-Meier curve between high and low risk groups. (E) Single factor Cox regression analysis forest map. (F) Multivariate Cox regression analysis forest map. ROC, receiver operating characteristic; AUC, area under the curve.

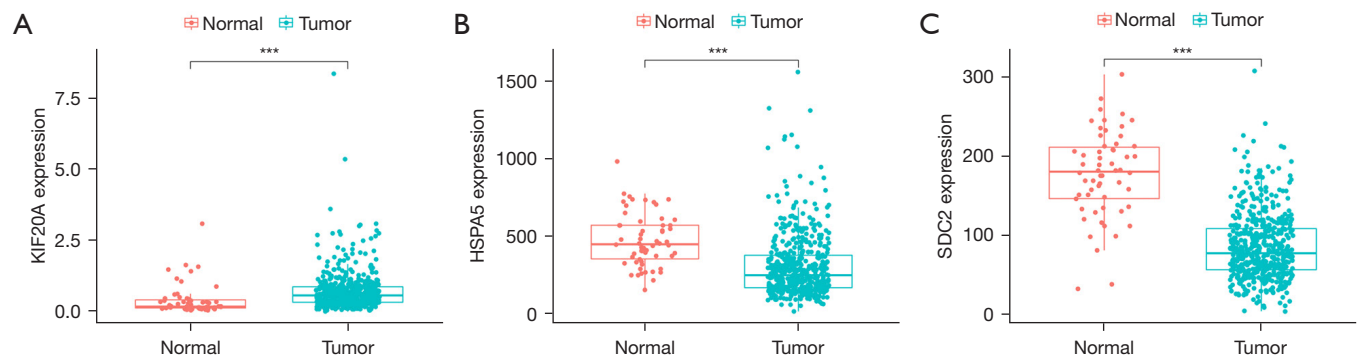


Figure 3 Expression of HSPA5, KIF20A and SDC2 in THCA and normal tissues. (A) Expression of KIF20A. (B) Expression of HSPA5. (C) Expression of SDC2. ***, $P < 0.001$. THCA, thyroid cancer.

ROC curve, 1.84 was the best cutoff point. The sensitivity is 0.67, and the specificity is 0.61 in this point. After evaluating the samples and using the 'predict ()' function to calculate the risk score based on the GRGs model, the samples were divided into the high-risk groups and the low-risk groups using the median (50th percentile cutoff point) risk score as the boundary. The expression of the three model genes in the high-risk and the low-risk groups is provided in *Figure 2B*.

Verification of GRGs model

The obtained genomic data were used to verify the GRGs model. The overall survival (OS) of the low-risk group was significantly higher than that of the high-risk group (*Figure 2C, 2D*); In the lower half of *Figure 2C*, the y-axis represents the survival time, the red represents the dead patients in this group, and the blue dot represents the living patients in this group. In the upper part of *Figure 2C*, the y-axis represents the risk score, the dotted line represents the grouping, the green dot represents the patients with low-risk, and the red represents the patients with high-risk. As can be seen from the upper and lower parts of composite *Figure 2C*, the number of deaths in the low-risk group was lower than that in the high-risk group. Prognosis was significantly affected by age, stage, and GRGs model according to the univariate risk regression analysis (*Figure 2E*). P value of age was 0.035, and the Hazard ratio (HR) of age was 1.02. P value of stage was less than 0.001, and the HR of age was 1.54. P value of the risk score was less than 0.001, and the HR of the risk score was 1.22. Only GRGs model was capable of being used as an independent prognostic factor in multivariate risk

regression analyses (*Figure 2F*). P value of the risk score was less than 0.001, and the HR of the risk score was 1.20. Leave-one-out analysis of the risk score cross validation showed the area under ROC curve (AUC) was 0.68.

Expression and mutation of model genes

An analysis of the expression levels of the genes was conducted. *KIF20A* (*Figure 3A*) was highly expressed while *HSPA5* (*Figure 3B*) and *SDC2* (*Figure 3C*) were decreased in THCA. *HSPA5*, *KIF20A* and *SDC2*'s expression is 477.80, 0.40, 177.72 respectively in normal samples, while expression of which is 307.02, 0.72, 86.77 respectively, in tumor samples. An analysis of the immunohistochemical images acquired from the HPA website of the three genes in normal thyroid gland tissues and THCA was conducted and this analysis confirmed these findings (17). The Human Protein Atlas version 21.0 provided the immunohistochemical images of the three genes shown in *Figure 4*. As a result of identification of mutations in model genes, *SDC2* and *KIF20A* exhibited lower mutation rates compared to *HSPA5* (*Figure S1*).

GRGs model and clinical characteristics

An analysis of clinical demographics and survival rates of THCA patients was conducted to discover their relationship. In the analysis, age ($P=0.10$) and gender ($P=0.14$) were not statistically significant (*Figure 5A, 5B*). At the same time, only grades (*Figure 5C*) and T stages (*Figure 5D*) significantly influenced survival rate of patients ($P < 0.001$).

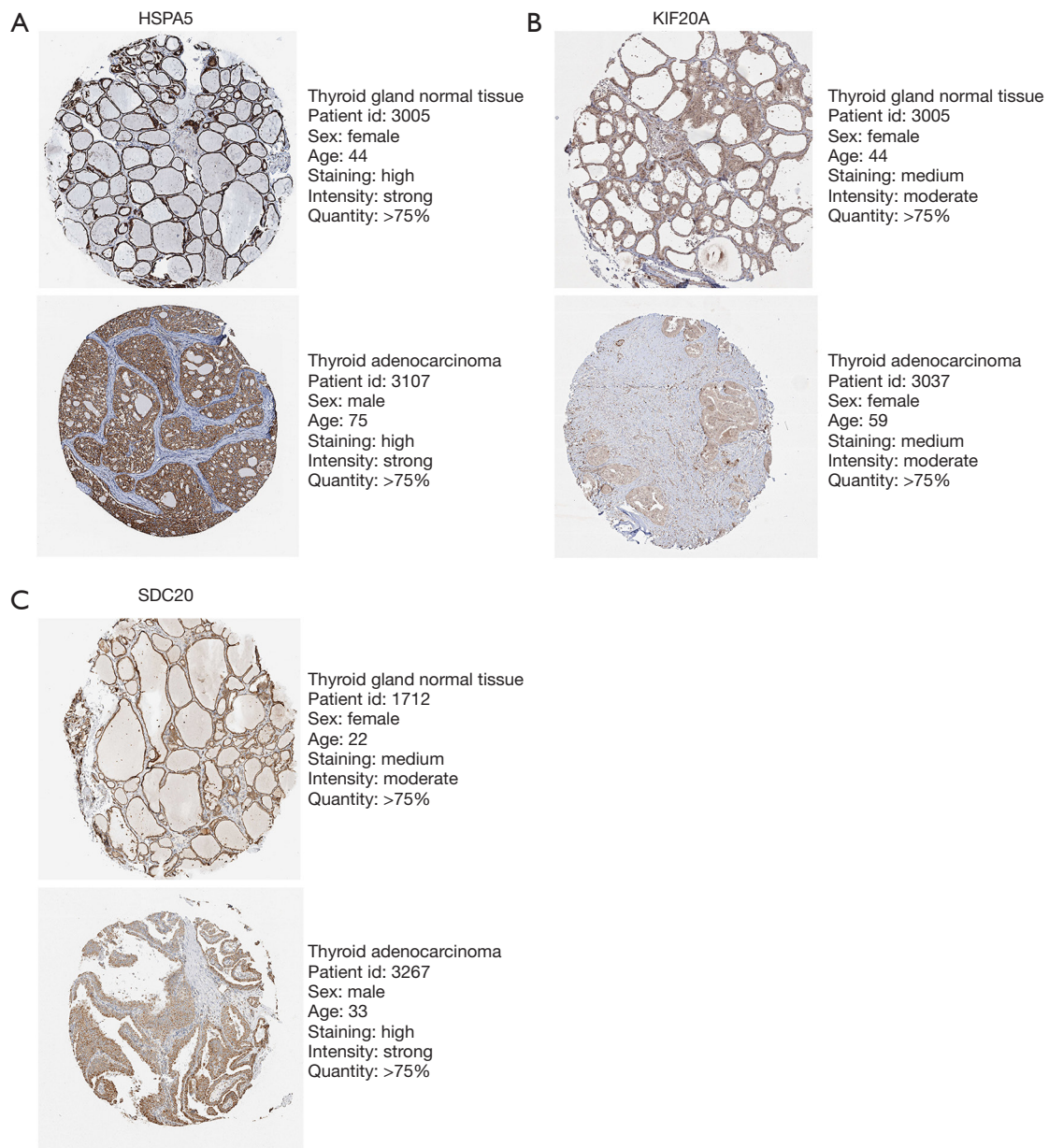


Figure 4 Validation of model GRG IHC images. (A) Validation of HSPA5 IHC images. (B) Validation of KIF20A IHC images. (C) Validation of SDC2 IHC images. Scale bar = 200 μ m. GRG, glycolysis-related gene; IHC, immunohistochemical.

Based on the GRGs model, the clinical traits were grouped and the survival rates were analyzed later. Six patients with varying clinical characteristics could be explicitly distinguished by GRGs ($P < 0.05$, *Figure 6*). This indicated that this GRGs model can distinguish high-risk and low-risk groups in the face of these six clinical characteristics. Nonetheless, result from distinguishing patients using age > 65 , male, M0–1, N0, Stage III–IV

(*Figure S2*) was not significant.

Discussion

THCA is the most common endocrine malignancy worldwide. In recent years, the incidence of THCA has been increasing (18,19). Although the incidence is high, the pathogenesis of thyroid has not been elucidated, so we need

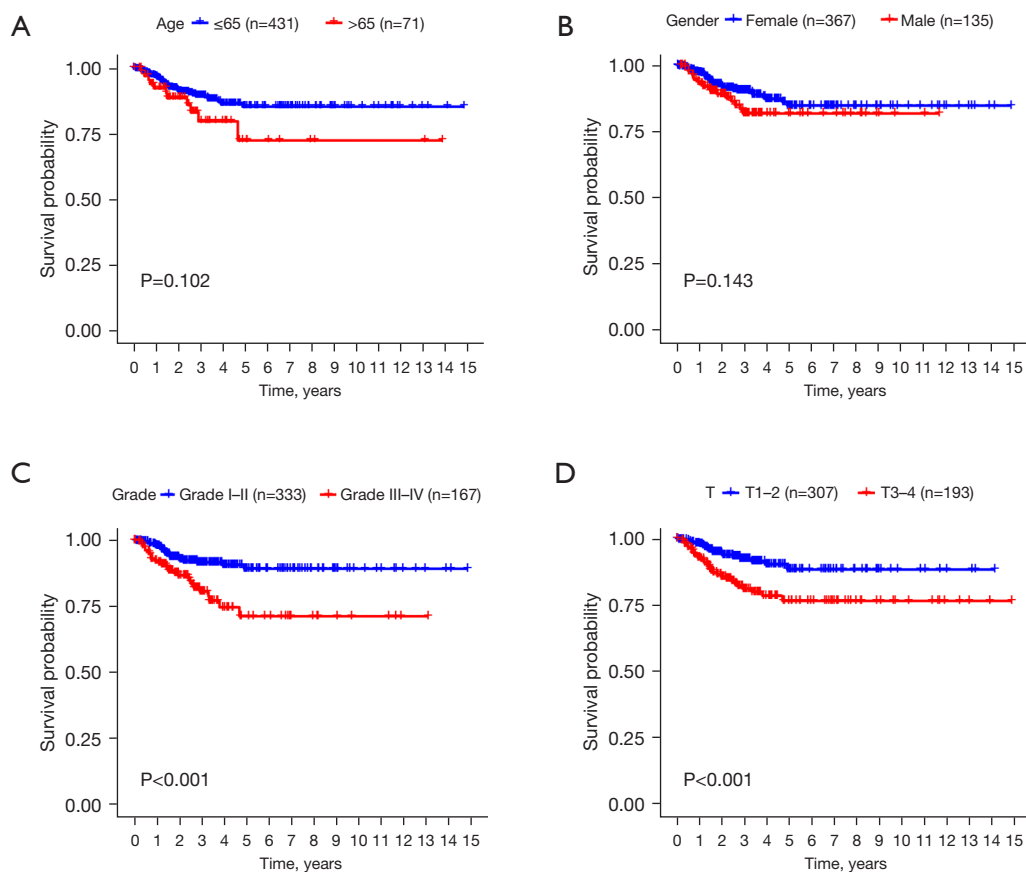


Figure 5 Relationship between different clinical characteristics and survival of patients with THCA. (A) Between age and survival of patients with THCA. (B) Between gender and survival of patients with THCA. (C) Between stage and survival of patients with THCA. (D) Between T stage and survival of patients with THCA. THCA, thyroid cancer.

to further study. The purpose of this study is to identify GRGs in the process of THCA progression and metastasis. In this study, we constructed a glycolytic gene model associated with the prognosis of THCA patients, including three genes (*HSPA5*, *KIF20A* and *SDC2*). Our model can serve as a significant independent prognostic factor for THCA. Through this model, we can effectively predict the tumor metastasis and prognosis of THCA patients, and better select personalized programs for further treatment of patients.

In clinical practice, thyroid nodules are very common (20). Ultrasound guided fine needle aspiration (US-FNA) is the most commonly available way to evaluate thyroid nodules. However, rapid On-Site Evaluation (ROSE) is considered to be the most helpful with small sized nodules or nodules that are more difficult to sample for less experienced radiologists (21). PTC patients with appropriate treatment and radiotherapy have a high 10-year

survival rate (22). In radiotherapy, we must consider the harmful effects of radiation for normal tissues surrounding the tumor tissues. Accurate calculation of out-of-field dose to be critical for informing risk estimates, such as estimation of out-of-field dose variation using Monte Carlo simulations (23). Therefore, exploring novel therapeutic targets of THCA is still a major challenging issue.

Recently, studies on cellular energy metabolism have attracted people's attention, especially the Warburg effect. Altered energy metabolism is a biochemical fingerprint of cancer cells that represents one of the "hallmarks of cancer". Even when oxygen is abundant, cancer cells count primarily on glycolysis to produce ATP energy unlike normal cells (24). There has also been a great deal of research done on THCA's glucose metabolism (25). In order to find probable biological targets for the treatment and diagnosis of THCA, the study of THCA glycolytic related genes and pathways is a potential approach. In addition, the current

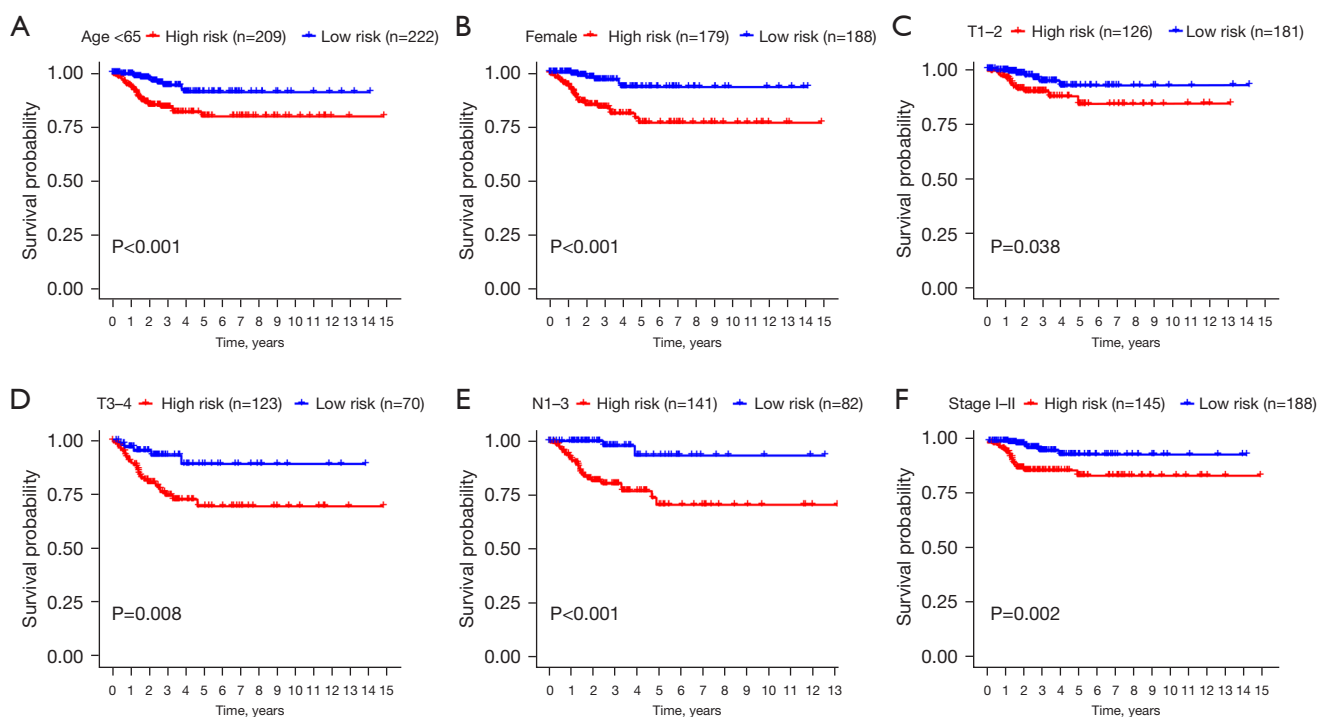


Figure 6 Relationship between clinical features and survival of patients with THCA. (A) Relationship between age <65 and survival of patients with THCA after model grouping. (B) Relationship between female and survival of patients with THCA after model grouping. (C) Relationship between T1–2 and survival of patients with THCA after model grouping. (D) Relationship between T3–4 and survival of patients with THCA after model grouping. (E) Relationship between N1–3 and survival of patients with THCA after model grouping. (F) Relationship between stage I–II and survival of patients with THCA after model grouping. THCA, thyroid cancer.

research on THCA glycolytic pathway mainly focuses on the role of single lncRNA in tumor growth and progression (26,27). It has also been proved that the expression level of glycolysis related proteins is different in different THCA subtypes, and is related to poor prognosis (28). This research is based on the protein molecular level. From the perspective of cell metabolism, this paper finds out the glycolytic genes related to THCA through the gene level. At the same time, as in this paper, we analyze the relationship between GRGs and survival rate in THCA patients by constructing a prediction model and sequencing clinical patient samples.

By analyzing the THCA data in TCGA, our study analyzed and found three glycolytic genes. A prognostic model, which is correlated with overall survival rate of THCA, was constructed using three GRGs (*HSPA5*, *KIF20A* and *SDC2*). Leave-one-out cross validation and ROC analysis indicated that the gene feature had a moderate effect on predicting survival rate. Previous studies suggested that the three GRGs we found were all

related to the occurrence and development of some cancers (29–31). For example, decreased expression of *HSPA5* is associated with the deterioration of clinical pathological characteristics, shortened PFI, and increased immune infiltration of various immune cells, indicating that *HSPA5* is a biomarker for the prognosis of THCA (32). It has also been pointed out that *HSPA5* contributes to the survival of head and neck cancer by maintaining lysosomal activity (33). These studies indicate that our results are consistent with those of our predecessors to a certain extent. Furthermore, multiple roles of *KIF20A* in central spindle assembly and cleavage furrow formation in mitosis and meiosis (34). *KIF20A* is highly expressed in many cancers, such as lung adenocarcinoma and hepatocellular carcinoma, and is a potential target for tumor gene therapy (35,36). *SDC2* methylation is a feasible biomarker for colorectal cancer detection (37). These results suggest that these glycolytic related genes may play an important role in determining the prognosis of patients with THCA.

We know that the current TMN clinical staging of

tumor is the most widely used malignant potential and the methods of disease progression. However, this method also has some disadvantages. For instance, in the case of patients suffering from the same disease at the same stage are treated with the same treatment, the prognosis of patients can make a big different. Despite being in the same stage, clinical outcomes differ for patients, indicating that the existing staging system is not sufficient for effective prognosis. Our study showed that only T stage was significantly associated with OS of THCA patients in the analyzed sample. In addition, there is no significant relationship between the survival rate of patients and the age and sex of cancer. Then, using the established model, we can identify high and low risk groups through six clinical characteristics. These six characteristics are age <65, female, T1–2, T3–4, N1–3, Stage I–III (Figure 6). However, we could not judge the risk by other characteristics in Figure S2. A novel method for assessing THCA patients is provided by our model, which can predict THCA patients' prognosis more accurately.

In addition, since our research is a retrospective analysis, there are some limitations. A prospective study using GRGs model for prospective research would be more persuasive. Moreover, in the analysis, the majority of the data was obtained from developed countries' public databases, but there was a lack of data from developing countries. It also needs to be reiterated that this is a secondary analysis of previous studies, and a prospective clinical study is needed to verify the accuracy of our results.

Conclusions

In conclusion, this study has identified a GRG model (including *HSPA5*, *KIF20A* and *SDC2*) which is related to the prognosis of THCA patients. Furthermore, these three GRGs may be potential targets for the treatment of THCA, while further prospective large-scale clinical trial should be conducted to confirm our conclusion.

Acknowledgments

Funding: This study was funded by the special project of "Research on Smart Education and Teaching Digital Transformation" in colleges and universities of Jiangsu Province (No. 2022ZHSZ60).

Footnote

Reporting Checklist: The authors have completed the

TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-22-2548/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-22-2548/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: 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. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Kitahara CM, Schneider AB. Epidemiology of Thyroid Cancer. *Cancer Epidemiol Biomarkers Prev* 2022;31:1284-97.
2. Deng Y, Li H, Wang M, et al. Global Burden of Thyroid Cancer From 1990 to 2017. *JAMA Netw Open* 2020;3:e208759.
3. Du L, Zhao Z, Zheng R, et al. Epidemiology of Thyroid Cancer: Incidence and Mortality in China, 2015. *Front Oncol* 2020;10:1702.
4. Wu M, Yuan H, Li X, et al. Identification of a Five-Gene Signature and Establishment of a Prognostic Nomogram to Predict Progression-Free Interval of Papillary Thyroid Carcinoma. *Front Endocrinol (Lausanne)* 2019;10:790.
5. Carling T, Udelsman R. Thyroid cancer. *Annu Rev Med* 2014;65:125-37.
6. Albero A, López JE, Torres A, et al. Effectiveness of chemotherapy in advanced differentiated thyroid cancer: a systematic review. *Endocr Relat Cancer* 2016;23:R71-84.
7. Udelsman R, Zhang Y. The epidemic of thyroid cancer in the United States: the role of endocrinologists and ultrasounds. *Thyroid* 2014;24:472-9.

8. Brito JP, Morris JC, Montori VM. Thyroid cancer: zealous imaging has increased detection and treatment of low risk tumours. *BMJ* 2013;347:f4706.
9. Kushchayeva Y, Kushchayev S, Jensen K, et al. Impaired Glucose Metabolism, Anti-Diabetes Medications, and Risk of Thyroid Cancer. *Cancers (Basel)* 2022;14:555.
10. Wen S, Luo Y, Wu W, et al. Identification of lipid metabolism-related genes as prognostic indicators in papillary thyroid cancer. *Acta Biochim Biophys Sin (Shanghai)* 2021;53:1579-89.
11. Jones NP, Schulze A. Targeting cancer metabolism--aiming at a tumour's sweet-spot. *Drug Discov Today* 2012;17:232-41.
12. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009;324:1029-33.
13. Kozal K, Józwiak P, Krześlak A. Contemporary Perspectives on the Warburg Effect Inhibition in Cancer Therapy. *Cancer Control* 2021;28:10732748211041243.
14. Iorio J, Petroni G, Duranti C, et al. Potassium and Sodium Channels and the Warburg Effect: Biophysical Regulation of Cancer Metabolism. *Bioelectricity* 2019;1:188-200.
15. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem Sci* 2016;41:211-8.
16. Schwartz L, Supuran CT, Alfarouk KO. The Warburg Effect and the Hallmarks of Cancer. *Anticancer Agents Med Chem* 2017;17:164-70.
17. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015;347:1260419.
18. La Vecchia C, Malvezzi M, Bosetti C, et al. Thyroid cancer mortality and incidence: a global overview. *Int J Cancer* 2015;136:2187-95.
19. Zha T, Wu H. Expression of serum AMPD1 in thyroid carcinoma and its clinical significance. *Exp Ther Med* 2018;15:3357-61.
20. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* 2016;26:1-133.
21. Medina Chamorro FM, Calle JA, Stein JE, et al. Experience of the Implementation of Rapid On-Site Evaluation in Ultrasound-Guided Fine-Needle Aspiration Biopsy of Thyroid Nodules. *Curr Probl Diagn Radiol* 2018;47:220-4.
22. Leboulleux S, Rubino C, Baudin E, et al. Prognostic factors for persistent or recurrent disease of papillary thyroid carcinoma with neck lymph node metastases and/or tumor extension beyond the thyroid capsule at initial diagnosis. *J Clin Endocrinol Metab* 2005;90:5723-9.
23. Sánchez-Nieto B, Medina-Ascanio KN, Rodríguez-Mongua JL, et al. Study of out-of-field dose in photon radiotherapy: A commercial treatment planning system versus measurements and Monte Carlo simulations. *Med Phys* 2020;47:4616-25.
24. Akram M. Mini-review on glycolysis and cancer. *J Cancer Educ* 2013;28:454-7.
25. Wu M, Ou-Yang DJ, Wei B, et al. A Prognostic Model of Differentiated Thyroid Cancer Based on Up-Regulated Glycolysis-Related Genes. *Front Endocrinol (Lausanne)* 2022;13:775278.
26. Jiang B, Chen Y, Xia F, et al. PTCSC3-mediated glycolysis suppresses thyroid cancer progression via interfering with PGK1 degradation. *J Cell Mol Med* 2021;25:8454-63.
27. Huo N, Cong R, Sun ZJ, et al. STAT3/LINC00671 axis regulates papillary thyroid tumor growth and metastasis via LDHA-mediated glycolysis. *Cell Death Dis* 2021;12:799.
28. Nahm JH, Kim HM, Koo JS. Glycolysis-related protein expression in thyroid cancer. *Tumour Biol* 2017;39:1010428317695922.
29. Zhang C, Liu Q, Zhou Y, et al. HSPA5 Could Be a Prognostic Biomarker Correlated with Immune Infiltration in Breast Cancer. *Dis Markers* 2022;2022:7177192.
30. Saito K, Ohta S, Kawakami Y, et al. Functional analysis of KIF20A, a potential immunotherapeutic target for glioma. *J Neurooncol* 2017;132:63-74.
31. Lei R, Zhao Y, Huang K, et al. The methylation of SDC2 and TFPI2 defined three methylator phenotypes of colorectal cancer. *BMC Gastroenterol* 2022;22:88.
32. Dong W, Du D, Huang H. HSPA5 is a prognostic biomarker correlated with immune infiltrates in thyroid carcinoma. *Endokrynol Pol* 2022;73:680-9.
33. Kim SY, Kim HJ, Kim HJ, et al. HSPA5 negatively regulates lysosomal activity through ubiquitination of MUL1 in head and neck cancer. *Autophagy* 2018;14:385-403.
34. Wu WD, Yu KW, Zhong N, et al. Roles and mechanisms of Kinesin-6 KIF20A in spindle organization during cell division. *Eur J Cell Biol* 2019;98:74-80.
35. Wu C, Qi X, Qiu Z, et al. Low expression of KIF20A suppresses cell proliferation, promotes chemosensitivity and is associated with better prognosis in HCC. *Aging*

- (Albany NY) 2021;13:22148-63.
36. Zhao X, Zhou LL, Li X, et al. Overexpression of KIF20A confers malignant phenotype of lung adenocarcinoma by promoting cell proliferation and inhibiting apoptosis. *Cancer Med* 2018;7:4678-89.
37. Zhang L, Dong L, Lu C, et al. Methylation of SDC2/TFPI2 and Its Diagnostic Value in Colorectal Tumorous Lesions. *Front Mol Biosci* 2021;8:706754.

Cite this article as: Wang B, Zhu Y, Zhang X, Wang Z. Identification and validation of a glycolysis-related gene signature for predicting metastasis and survival rate in patients with thyroid cancer. *Transl Cancer Res* 2023;12(5):1100-1111. doi: 10.21037/tcr-22-2548

Table S1 Fifty-two differentially expressed glycolysis-related genes

Gene	conMean	treatMean	logFC	P value
<i>AC016586.1</i>	0.541199	0.252303	-1.101	4.06E-14
<i>ACTN3</i>	0.05002	0.015548	-1.68578	1.12E-15
<i>AGRN</i>	22.1092	46.1904	1.062946	5.30E-27
<i>ANGPTL4</i>	8.47078	22.02692	1.378701	1.44E-10
<i>B3GAT1</i>	1.056808	4.091612	1.952957	1.29E-23
<i>B3GNT3</i>	0.343202	18.62594	5.762113	1.45E-18
<i>CD44</i>	23.64155	55.27598	1.225328	8.79E-24
<i>CHST2</i>	5.880716	25.84026	2.135557	1.19E-25
<i>CHST6</i>	0.036072	0.21004	2.541727	1.62E-17
<i>CITED2</i>	221.0603	68.17017	-1.69723	2.54E-26
<i>CLDN9</i>	1.030933	6.584371	2.675096	3.18E-20
<i>COL5A1</i>	2.005016	7.003881	1.804541	9.28E-08
<i>CTH</i>	5.16804	1.626699	-1.66767	3.30E-26
<i>DCN</i>	33.77995	8.477774	-1.99441	7.23E-25
<i>DPYSL4</i>	0.31194	0.944202	1.597825	0.037904
<i>DSC2</i>	1.184663	2.625639	1.148192	5.75E-08
<i>EFNA3</i>	0.706924	1.521173	1.105558	3.36E-20
<i>ELF3</i>	1.867485	7.641938	2.032842	1.70E-14
<i>ENO4</i>	0.437242	0.171567	-1.34966	1.14E-25
<i>FBP2</i>	0.098322	0.021292	-2.20719	9.49E-10
<i>GALE</i>	2.196281	18.68538	3.088775	8.78E-29
<i>GCKR</i>	0.02943	0.157077	2.416115	4.27E-14
<i>GPC3</i>	7.262081	0.815459	-3.1547	2.32E-32
<i>GPR87</i>	0.024971	0.420158	4.072619	2.33E-13
<i>GYS2</i>	0.096623	0.033104	-1.54537	9.34E-20
<i>HDAC4</i>	2.713549	1.10965	-1.29008	6.08E-27
<i>HKDC1</i>	4.254272	1.696967	-1.32595	4.16E-20
<i>HS6ST2</i>	0.321417	3.755313	3.546417	2.90E-07
<i>HTR2A</i>	0.050591	0.018263	-1.46996	2.90E-15
<i>IGFBP3</i>	8.277112	26.33518	1.669792	1.55E-10
<i>INSR</i>	6.885022	14.03193	1.027181	5.50E-26
<i>KDEL3</i>	4.718449	12.77691	1.437154	9.74E-23
<i>MET</i>	11.81067	69.16071	2.549862	4.53E-24
<i>MIOX</i>	9.232944	3.455103	-1.41806	8.10E-23
<i>MLXIPL</i>	0.744173	2.654179	1.834555	4.71E-14
<i>NT5E</i>	9.665598	26.37453	1.448214	7.61E-11
<i>OGDHL</i>	15.59248	5.8479	-1.41486	5.80E-22
<i>P4HA2</i>	2.849222	10.59953	1.895361	6.93E-32
<i>PC</i>	4.044894	11.66517	1.528034	1.09E-24
<i>PFKFB1</i>	0.405589	0.143794	-1.49602	2.64E-25
<i>PFKFB2</i>	18.7996	8.333122	-1.17377	1.99E-23
<i>PKP2</i>	0.244379	0.829063	1.762359	1.52E-09
<i>PPARGC1A</i>	8.456713	2.259498	-1.90409	1.31E-26
<i>QSOX1</i>	6.222757	18.44137	1.56732	3.62E-22
<i>SDC1</i>	22.69221	46.84309	1.045639	1.03E-17
<i>SDC2</i>	177.716	86.774	-1.03424	1.56E-23
<i>SDC3</i>	7.684957	19.25227	1.32492	5.21E-22
<i>TFF3</i>	802.1512	86.20233	-3.21808	6.66E-32
<i>TGFA</i>	3.401024	19.52026	2.520931	9.18E-28
<i>TGFBI</i>	4.196403	14.60663	1.799399	2.39E-17
<i>TKTL1</i>	0.029211	0.014387	-1.02176	2.49E-08
<i>VCAN</i>	0.912372	3.25451	1.834746	8.05E-05

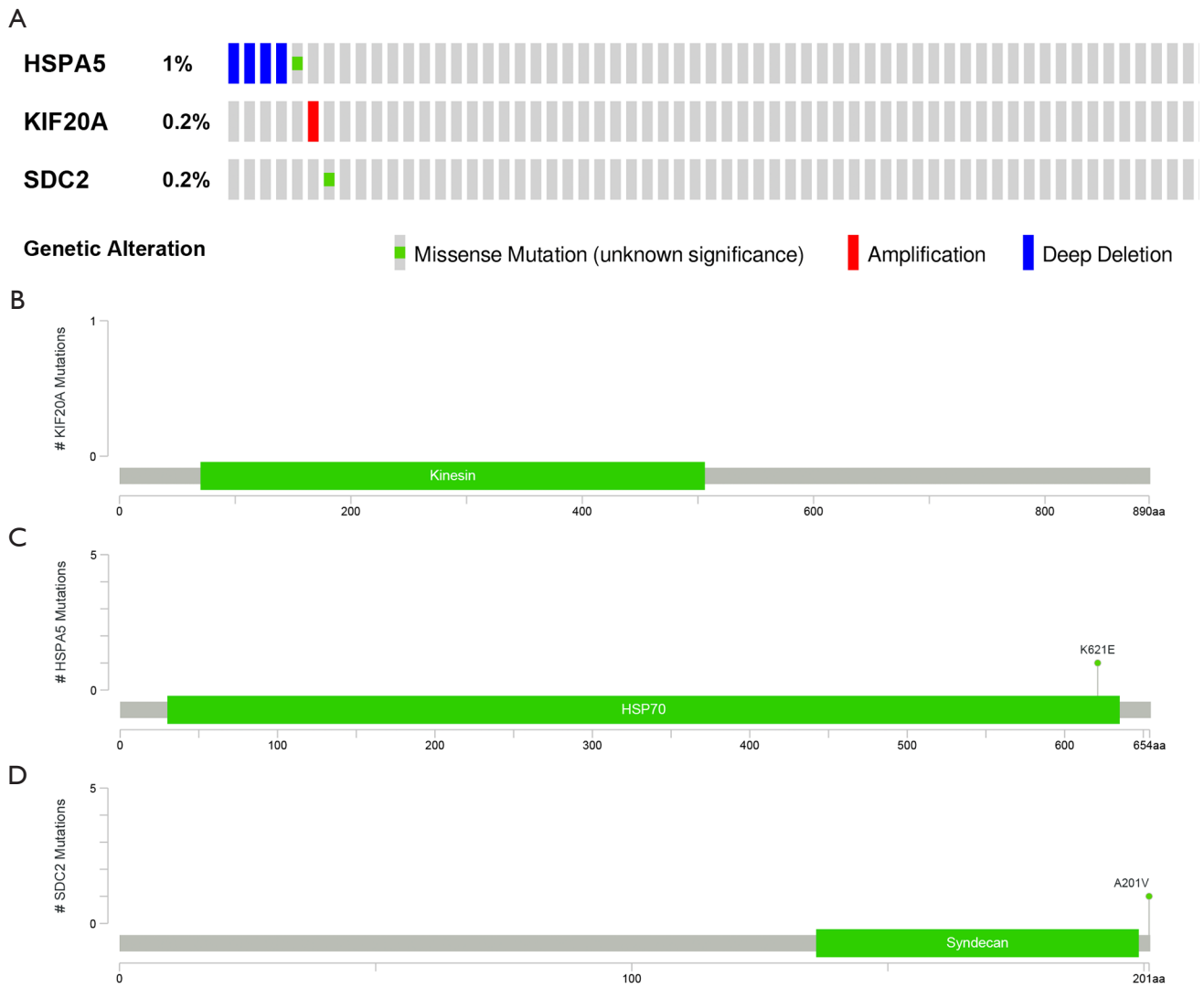


Figure S1 Identification of mutations in model genes.

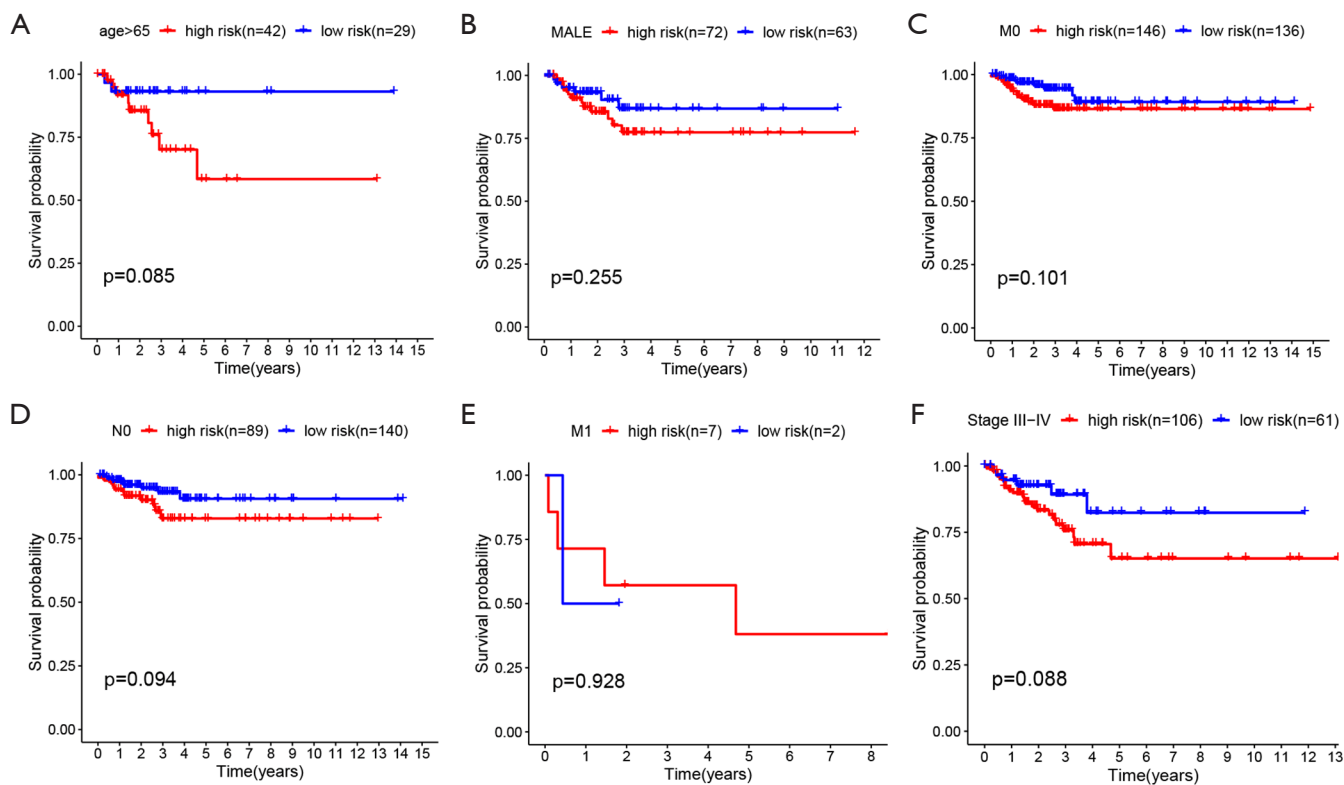


Figure S2 Relationship between survival probability and clinical features of thyroid cancer patients ($P>0.05$).