



A Four–Gene–Based Risk Score With High Prognostic Value in Gastric Cancer

Bingdong Zhang¹, Yuerui Li², Liu Yang³ and Yongbing Chen^{1*}

¹ Department of Gastrointestinal Surgery & Department of Clinical Nutrition, Beijing Shijitan Hospital, Capital Medical University, Beijing, China, ² Geriatric Cardiology Department of The Second Medical Center & National Clinical Research Center for Geriatric Diseases, Chinese People's Liberation Army of China General Hospital, Beijing, China, ³ State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Medicine, Tsinghua University, Beijing, China

Background: Gastric adenocarcinoma is an important contributor to cancer mortality and morbidity. This study aimed to explore the prognostic value of mutation patterns in gastric adenocarcinoma.

Materials and Methods: We extracted somatic mutation data for 437 gastric adenocarcinoma samples from The Cancer Genome Atlas (TCGA) Stomach Adenocarcinoma (STAD) cohort. Kaplan–Meier survival in the R package maftools was used to analyze associations between mutations and survival. Multivariate Cox proportional model was used to establish risk formula. A four-gene-based risk score was developed to predict the overall survival of patients with gastric adenocarcinoma. We used the Tianjin cohort dataset with survival information to further evaluate the clinical value of this mutation signature.

Results: Forty-five survival-related mutated genes were identified and verified, most of which were co-occurring in their mutation pattern and co-occurring with MLH3 and polymerase ϵ (POLE) mutations. Gastric adenocarcinoma samples with the 45 mutated genes had a significantly higher mutation count. Four-gene [UTRN, MUC16, coiled-coil domain-containing protein 178 (CCDC178), and HYDIN] mutation status was used to build a prognostic risk score that could be translated into the clinical setting. The association between the four-gene-based signature and overall survival remained statistically significant after controlling for age, sex, TNM stage, and POLE mutation status in the multivariate model [hazard ratio (HR), 1.88; 95% CI, 1.33–2.7; $p < 0.001$]. The prognostic significance of the four-gene-based risk score identified in TCGA cohort was validated in the Tianjin cohort.

Conclusion: A four-mutated gene risk formula was developed that correlated with the overall survival of patients with gastric adenocarcinoma using a multivariable Cox regression model. In two independent genomic datasets from TCGA and Tianjin cohorts, low risk scores were associated with higher tumor mutation loads and improved outcome in patients with gastric adenocarcinoma. This finding may have implications for prognostic prediction and therapeutic guidance for gastric adenocarcinoma.

Keywords: gastric cancer, risk score, mutations, prognostic value, genome

OPEN ACCESS

Edited by:

Fenglin Liu,
Fudan University, China

Reviewed by:

Michele Ghidini,
IRCCS Foundation Ca' Granda
Ospedale Maggiore Policlinico, Italy
Yiqi Du,
Second Military Medical University,
China

*Correspondence:

Yongbing Chen
chyb197404@163.com

Specialty section:

This article was submitted to
Gastrointestinal Cancers,
a section of the journal
Frontiers in Oncology

Received: 16 July 2020

Accepted: 06 August 2021

Published: 02 September 2021

Citation:

Zhang B, Li Y, Yang L and Chen Y
(2021) A Four–Gene–Based Risk
Score With High Prognostic
Value in Gastric Cancer.
Front. Oncol. 11:584213.
doi: 10.3389/fonc.2021.584213

INTRODUCTION

Gastric adenocarcinoma is an important contributor to cancer mortality and morbidity, and its molecular mechanism remains largely incomprehensible. Next-generation sequencing (NGS) technology could provide genomic-level information about the mechanism of cancer. Numbers of large-scale genomic analyses on gastric adenocarcinoma have been completed, including The Cancer Genome Atlas (TCGA) project.

Recent research has shown that gastric adenocarcinoma is a heterogeneous disease. Surgical resection is still the main means of curative treatment for gastric adenocarcinoma. However, a portion of patients with advanced gastric adenocarcinoma developed local recurrences and distant metastases and had a poor prognosis after resection (1). Patients would have received the best treatment if their prognosis was depicted in advance. However, different prognoses of patients with similar clinical stages or pathologic grades remain unpredictable (2–5). Profiling the genetic mutation of gastric adenocarcinoma that influences the prognosis and accurate risk assessment based on genetic screening will lead to more effective clinical strategies in precision medicine.

In this study, we identified and verified 45 survival-related mutated genes with bioinformatics analysis from TCGA Stomach Adenocarcinoma (STAD) cohort. We investigated the function of these genes *via* Gene Ontology (GO) analysis. Through random survival forest algorithm, we ranked these mutated genes by importance and constructed a four-gene-based risk score with multivariable Cox regression model. Using the Tianjin cohort dataset with survival information, we evaluated the clinical value of the risk score.

MATERIALS AND METHODS

Stomach Adenocarcinoma Datasets

Genomic data of gastric adenocarcinoma somatic mutation and gene expression data for 437 gastric adenocarcinoma samples in TCGA data portal (level 3) were downloaded from Genomic Data Commons (<https://portal.gdc.cancer.gov>) (Release Date: August 23, 2018). The Tianjin cohort contained data from 78 patients from northern China (6). Frozen tissue samples derived from surgical resection specimens of primary gastric adenocarcinoma from 294 northern Chinese patients without preoperative chemotherapy or radiotherapy were obtained from the Tianjin Medical University Cancer Institute and Hospital-National Foundation for Cancer Research Joint Tissue Banking Facility. Whole-exome sequencing was performed on 78 samples. Germline DNA was obtained from matching blood samples and used as a reference sequence to detect somatic mutations. Histopathologic diagnoses were independently reviewed by at least two experienced pathologists. Clinical follow-up data were complete for 78 participants with 25.08 months of median follow-up (32 deceased, 41.03%) (**Supplementary Table 1**). TCGA cohort also has the follow-up and vital status of patients (**Supplementary Table 2**). This study was approved by the Chinese PLA General Hospital (Beijing), which waived additional informed consent because all data used in this study were obtained from

public databases. This study met the publication guidelines provided by TCGA (<http://cancergenome.nih.gov/publications/publicationguidelines>). All data were processed and analyzed by Excel 2010 and R (version 3.5.0).

Prognosis

Kaplan–Meier survival analyses implemented in the R package *maftools* and survival were used to analyze the correlation between mutations and survival (7). The log-rank test was used to determine significant differences of survival curves stratified by mutations. A two-sided $p < 0.05$ was considered statistically significant. Correlation between mutations and survival was also explored by multivariate Cox regression analyses by the R package *survival*.

Gene Ontology Pathway Analysis

The GO pathway analysis mutated genes were annotated by the R package of *clusterProfiler* (8). The cutoff p -adjust value was 0.01.

Random Survival Forest Algorithm

Random survival forest algorithm implemented in the R *randomForestSRC* was used to rank the survival-related genes by their importance.

RESULTS

Survival in The Cancer Genome Atlas Stomach Adenocarcinoma Cohort

Somatic mutation data for 437 gastric adenocarcinoma samples from TCGA STAD cohort were extracted by *maftools* version 1.6.15 (<http://www.bioconductor.org/packages/release/bioc/vignettes/maftools/inst/doc/maftools.html>). These somatic mutations included point mutations and insertions/deletions (indels). We annotated 17,431 protein-coding genes with somatic mutations that mostly consist of missense mutations (**Figure 1A**), and single-nucleotide polymorphism (SNP) was more common than insertion–deletion (InDel) (**Figure 1B**). Furthermore, the mutational contexts are derived from combinations of six mutational types (i.e., T>G, T>C, T>A, C>T, C>G, and C>A) (**Figure 1C**). The somatic mutation rates varied considerably among the samples (**Figures 1D, E**), though an average of 108 mutations occurred in each sample (**Figure 1D**). The top 10 mutated genes were *TTN*, *TP53*, *MUC16*, *LRP1B*, *SYNE1*, *ARID1A*, *CSMD3*, *FAT4*, *FLG*, and *PCLO* (**Figure 1F**).

In order to select the most weighted genes associated with survival outcome, Kaplan–Meier survival in the R package *maftools* was used to analyze associations between mutations and survival for each of the 17,431 protein-coding genes (**Figure 1B**). A two-sided $p < 0.05$ was considered statistically significant. The most studied tumor repressor gene *TP53* was analyzed for example. All samples were categorized into two groups representing the wild-type and mutated *TP53* respectively, and Kaplan–Meier survival analysis showed that there were no significant associations between *TP53* mutations and survival in TCGA STAD cohort (**Figure 1G**). Forty-four

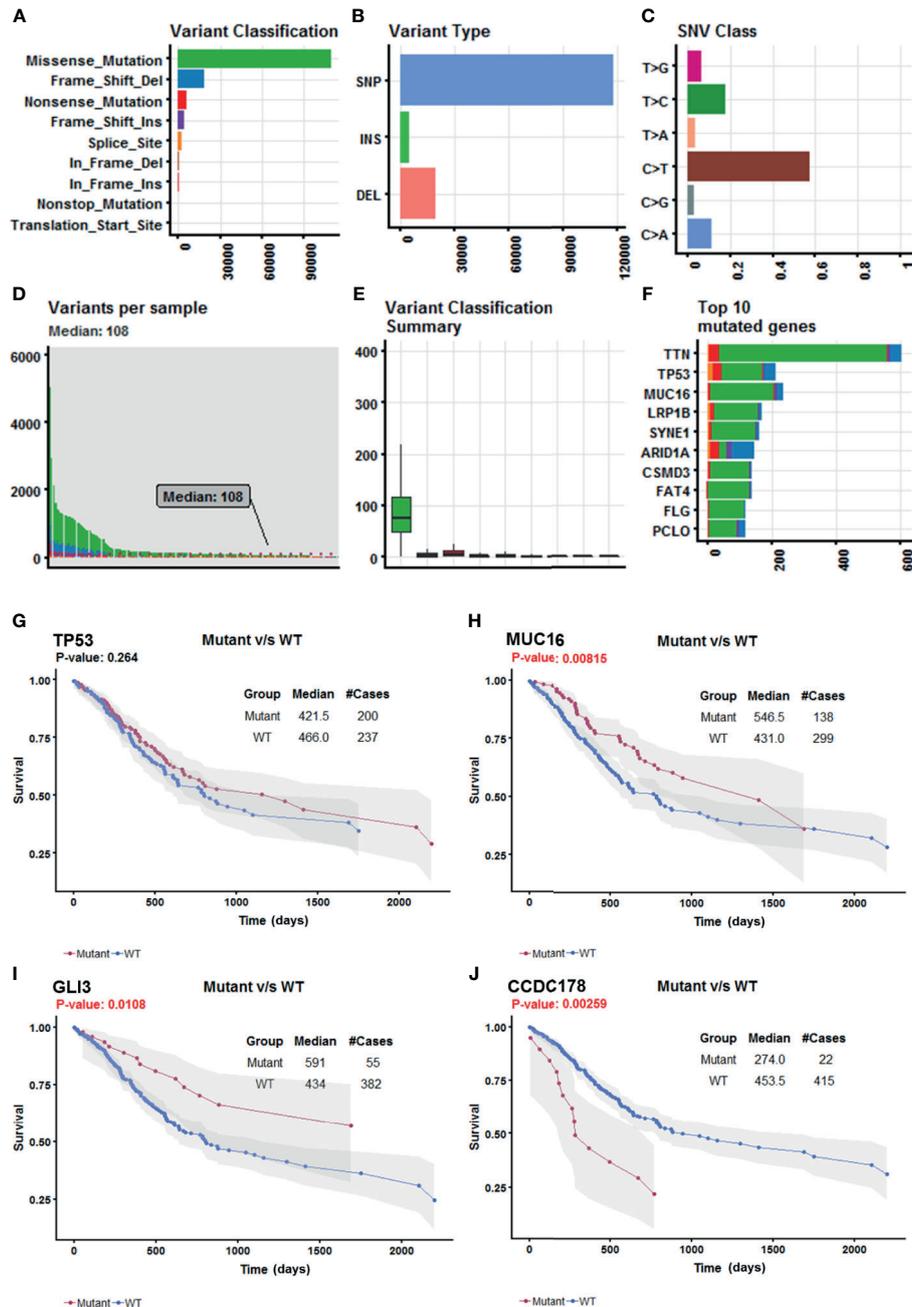


FIGURE 1 | Selection of survival-related mutated genes in The Cancer Genome Atlas (TCGA) Stomach Adenocarcinoma (STAD) cohort. **(A–C)** Characteristics of the variations in TCGA STAD cohort. Histogram summarizing the variant types of all cases with substitutions, insertions, deletions, and SNP (single nucleotide polymorphism). **(D)** Stacked bar showing the cumulative frequency of variation for individual cases. **(E)** Boxplot summarizing the number of cases for each type of variant. **(F)** Stacked bar showing top 10 mutated genes. **(G–J)** Kaplan–Meier survival analyses stratified by TP53, MUC16, GLI3, and coiled-coil domain-containing protein 178 (CCDC178) mutation status, respectively. This analysis was implemented in the R package maftools.

mutated genes, such as MUC16 and GLI3, occurring in more than 5% of the patients, were significantly associated with a better survival outcome (Figures 1H, I). Only coiled-coil domain-containing protein 178 (CCDC178) mutations were significantly associated with a poor survival outcome (Figure 1J). The 45 mutated genes were listed in Table 1.

To elucidate the function of these survival-related mutated genes, we conducted GO analysis and revealed that many genes play an important role in “cell–cell adhesion *via* plasma-membrane adhesion”, “extracellular matrix component”, and “alpha-catenin binding”, which were highly correlated to cancer metastasis and invasiveness (Table 2).

TABLE 1 | Survival-related mutated genes by Kaplan–Meier survival analyses.

Gene (Gene Symbol)	Median time in days	Median time in days	Mutated samples	Log-rank p-value
MUC16	546.5	431	138	0.00815
CSMD1	595	424	79	0.03983
GLI3	591	434	55	0.0108
PTPRT	503	436	50	0.02372
CDH23	600	429.5	48	0.00794
COL11A1	549	434	46	0.01143
MACF1	600	433.5	44	0.03886
CELSR3	679	426.5	43	0.00691
TENM3	485	439.5	43	0.02574
DCLK1	594	435	43	0.04194
TECTA	657.5	434	42	0.01237
ATM	688	428	39	0.02031
ITPR3	664	426.5	38	0.01208
ZNF462	762	422	37	0.02954
UTRN	522	439	36	0.01643
RANBP2	544	435	36	0.02873
HERC1	589.5	427	35	0.0406
FBXW7	644	434	35	0.0493
CHD6	569	434	32	0.0354
CPAMD8	669	427.5	31	0.03831
ANK1	639	434	30	0.01384
HYDIN	468	443	30	0.03202
CTNNB1	742	427.5	29	0.00461
ST8SIA6	695.5	427	29	0.04623
PCDH20	766	427.5	28	0.01026
KIF26B	640	431	28	0.03642
RAI1	670	428	28	0.04695
YLPM1	713	436	26	0.00804
PARD3B	809	431	26	0.01547
GREB1	805	432.5	26	0.02778
SOGA3	616	437.5	26	0.032
COL15A1	588	437.5	25	0.02058
ABCC8	812	427.5	24	0.02221
MYH14	668	440	24	0.0243
KIAA1549	678.5	431	24	0.04694
CDHR2	900	429.5	23	0.00268
MYH11	491	439.5	23	0.02104
PCDHB6	600	414	23	0.02661
COL5A2	900	432.5	23	0.04145
CCDC178	274	453.5	22	0.00259
ZZEF1	679	429.5	22	0.00588
THSD1	514.5	440	22	0.01973
DLGAP2	655.5	428	22	0.02265
OBSL1	616	435	22	0.03932
LAMB3	736.5	440	22	0.04177

Clinical Features of Patients in The Cancer Genome Atlas Stomach Adenocarcinoma Cohort With the 45 Mutated Genes

To explore the relationship between these survival-related mutations, we performed pair-wise Fisher's exact test to detect significant pairs of mutated genes in the 45 genes and DNA mismatch repair (MMR)-related genes [PMS2, MSH2, MLH1, MSH3, MLH3, MSH6, polymerase ϵ (POLE)]. Interestingly, most of the 45 mutated genes were co-occurring in their mutation pattern and co-occurring with MLH3 and POLE mutations (**Figure 2A**). Tumor mutation burden (TMB) is an important determinant for molecular subtyping of gastric adenocarcinoma in TCGA (9). Recent studies have shown that gastric adenocarcinoma with POLE mutations or microsatellite

instability-high (MSI-H) had DNA MMR signatures and higher TMBs (10). Gastric adenocarcinoma samples with the 45 mutated genes had a significantly higher tumor count (**Figure 2B**; Mann-Whitney test p-value <0.0001). Since these mutations tended to occur simultaneously, we explored the relationship between all these mutations and prognosis. Kaplan–Meier survival analysis showed that the 45 mutated genes were significantly associated with a better survival outcome in TCGA STAD cohort (**Figure 2C**; log-rank test, $p < 0.0001$). Multivariate Cox regression analyses showed that the correlation remained statistically significant after controlling for confounding factors such as sex, age, and TNM stage (**Figure 2D**). So, the 45 mutated genes were important prognostic indicators associated with tumor mutation load.

TABLE 2 | GO analysis of 45 survival-related mutated genes (partial data).

Genes	p-value	p.adjust	Annotations
Biological process			
7	7.50E-08	0.000105	GO:0007156:homophilic cell adhesion via plasma membrane adhesion molecules
7	1.92E-06	0.001342	GO:0098742:cell-cell adhesion via plasma-membrane adhesion molecules
Molecular function			
5	5.89E-05	0.0048705	GO:0051015:actin filament binding
3	7.38E-05	0.0048705	GO:0030020:extracellular matrix structural constituent conferring tensile strength
Cellular component			
4	3.91E-06	0.000673	GO:0044420:extracellular matrix component

Partial data, genes involved p.adjust < 0.01 (GO analysis).
GO, Gene Ontology.

Construction of Risk Score Formula

The most common mutations of the 45 genes are missense substitutions. Other alterations include silent mutations, frameshift insertions and deletions, nonsense mutations, and other infrequent mutations. Both base substitution mutation and frameshift mutation can change the composition or sequence of amino acids in the polypeptide chain. According to the impact of different type mutations on DNA composition or

sequence of amino acids, we scored these mutations, as follows: No mutation or Silent mutation: 0; Missense mutation, In-frame insertion or deletion: +1; Splice site mutation: +2; Nonsense mutation, frameshift insertion or deletion: +3; Multiple mutation: the maximum score of all (Figure 3A). There are four levels of mutations in total. Finally, we get the mutation score matrix (Supplementary Table 3). In order to select the most weighted genes, we used random survival forest algorithm

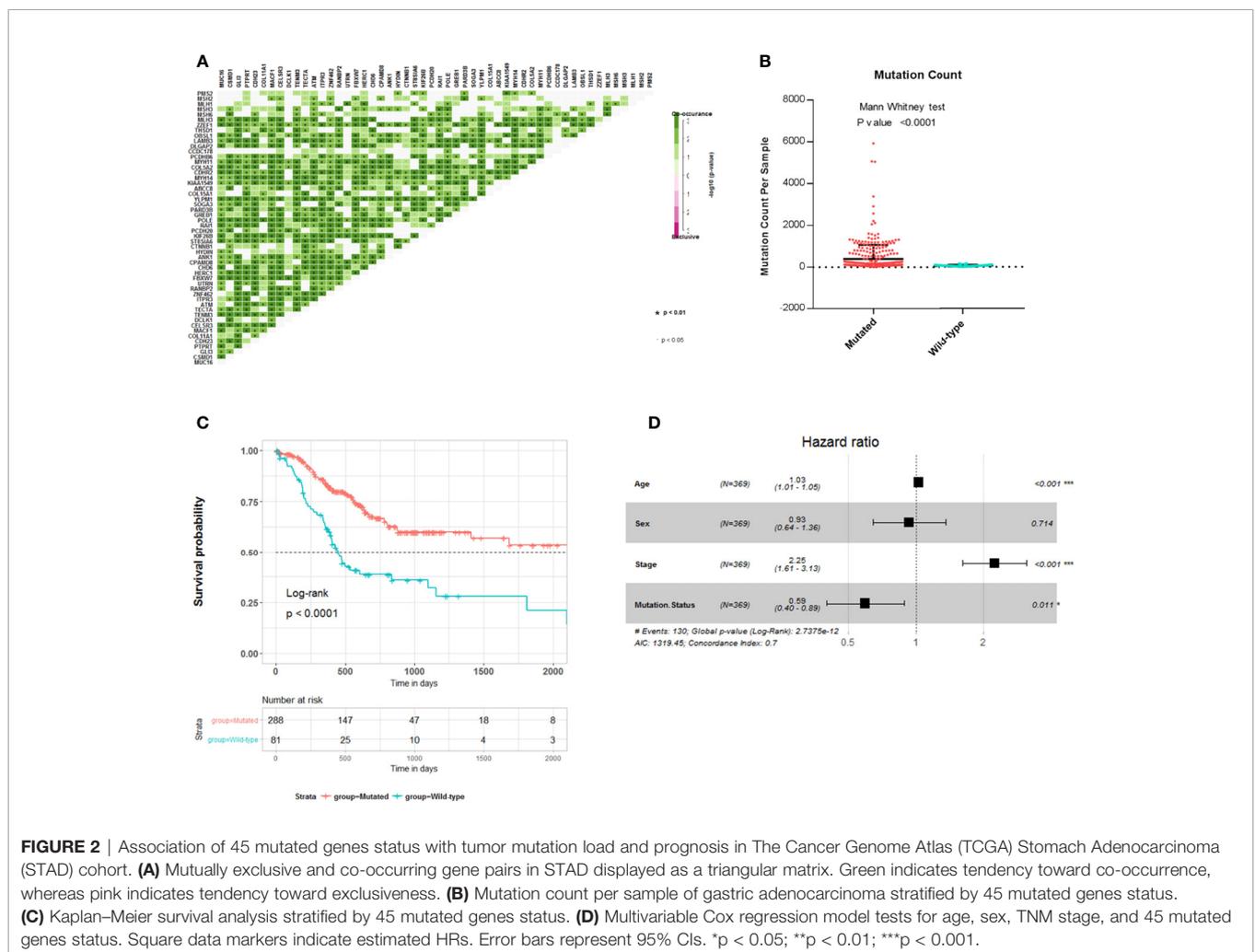


FIGURE 2 | Association of 45 mutated genes status with tumor mutation load and prognosis in The Cancer Genome Atlas (TCGA) Stomach Adenocarcinoma (STAD) cohort. **(A)** Mutually exclusive and co-occurring gene pairs in STAD displayed as a triangular matrix. Green indicates tendency toward co-occurrence, whereas pink indicates tendency toward exclusiveness. **(B)** Mutation count per sample of gastric adenocarcinoma stratified by 45 mutated genes status. **(C)** Kaplan–Meier survival analysis stratified by 45 mutated genes status. **(D)** Multivariable Cox regression model tests for age, sex, TNM stage, and 45 mutated genes status. Square data markers indicate estimated HRs. Error bars represent 95% CIs. *p < 0.05; **p < 0.01; ***p < 0.001.

by log rank). The correlation between the four-gene-based signature and overall survival remained statistically significant after controlling for age, sex, TNM stage, and POLE mutation status in the multivariate model [hazard ratio (HR), 1.88; 95% CI, 1.33–2.7; $p < 0.001$] (Figure 3G). A significantly higher mutation count was also observed in gastric adenocarcinoma samples within the low-risk group (Figure 3H; Mann–Whitney test p -value < 0.0001). So, the signature of the four-gene mutation would be a good prediction for survival of gastric adenocarcinoma patients.

Independent Validation of Four–Gene–Based Risk Score in the Tianjin Cohort

To further evaluate the clinical value of this four-gene mutation signature, we used Tianjin cohort dataset with survival information. Kaplan–Meier survival analyses showed low-risk scores were significantly associated with better survival outcomes (Figure 4A; log-rank test, $p = 0.036$). Significantly higher mutation count was also observed in Tianjin cohort gastric adenocarcinoma samples in the low-risk group (Figure 4B;

Mann–Whitney test p -value < 0.0001). Multivariable Cox regression analysis also showed that the association of the risk score with overall survival was statistically significant after controlling for age, sex, TNM stage, and POLE mutation status (Figure 4C; HR, 4.33; 95% CI, 1.29–14.5; $p = 0.018$).

DISCUSSION

We analyzed 437 gastric adenocarcinoma samples from TCGA cohort and 78 gastric adenocarcinoma samples from a Tianjin cohort for survival prediction genes. We have identified and verified 45 mutated genes related to survival from 17,431 protein-coding genes in STAD. The GO enrichment showed that these genes play an important role in “cell–cell adhesion *via* plasma–membrane adhesion”, “extracellular matrix component”, and “alpha-catenin binding”. Obviously, these genes played a pivotal role in cancer metastasis. After that, we ranked the 45 survival–related mutated genes by random survival forest algorithm. Whole-genome sequencing is expensive, and

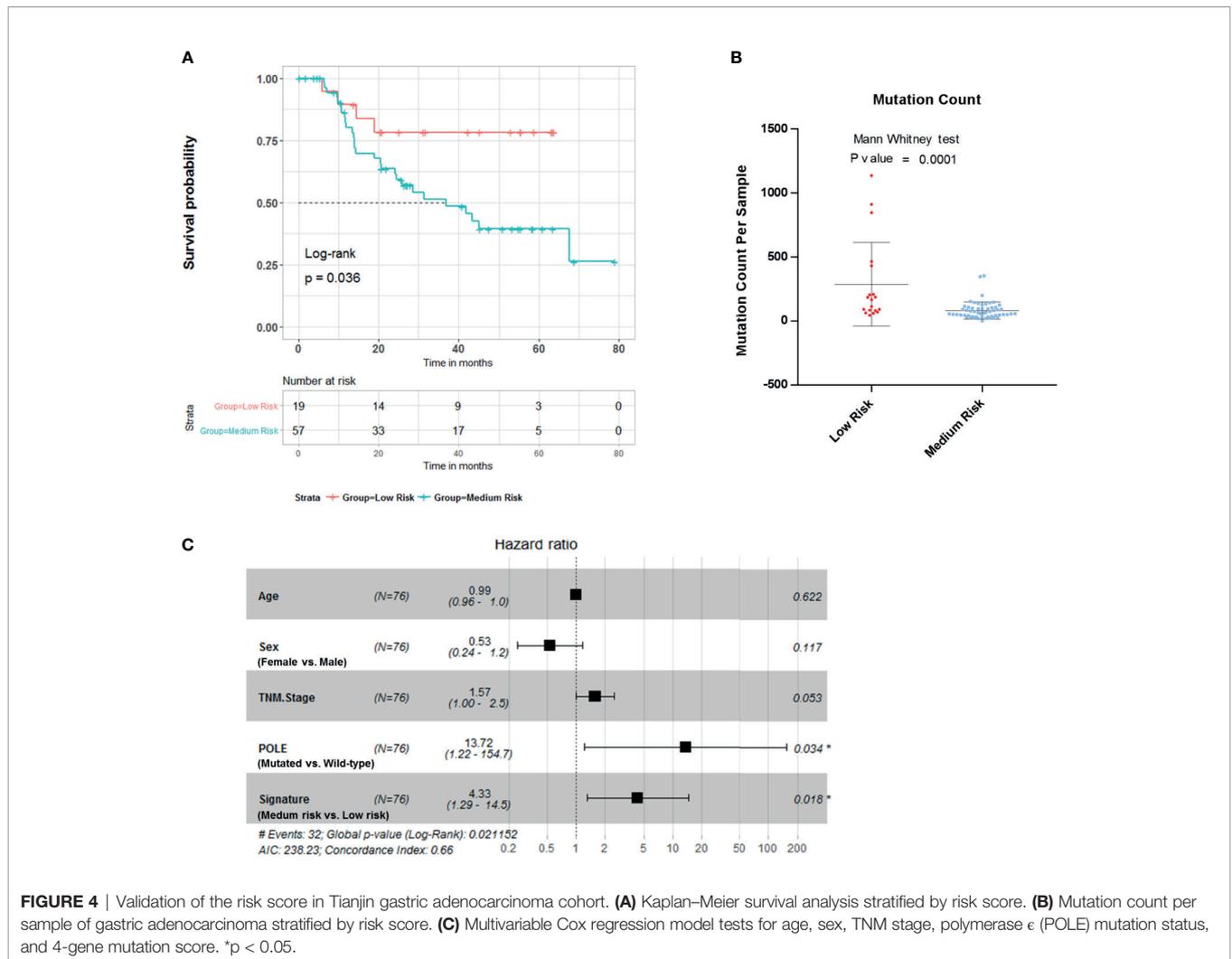


FIGURE 4 | Validation of the risk score in Tianjin gastric adenocarcinoma cohort. **(A)** Kaplan–Meier survival analysis stratified by risk score. **(B)** Mutation count per sample of gastric adenocarcinoma stratified by risk score. **(C)** Multivariable Cox regression model tests for age, sex, TNM stage, polymerase ϵ (POLE) mutation status, and 4-gene mutation score. * $p < 0.05$.

data analysis takes a long time. The aim of this study is to develop a cheap and practical prognostic tool that can be accomplished by PCR. Combining Kaplan–Meier survival analysis log-rank p-value, mutation coexistence pattern, and the result of random survival forest algorithm, we selected four-gene mutation status to build a prognostic risk score that could be transformed into the clinical setting. Gastric adenocarcinoma samples classified into low-risk group had a significantly higher tumor mutation load and better survival outcome. The association between the four-gene-based risk score and overall survival was independent of mutations in POLE mutation status, age, sex, and TNM stage. Many studies have reported the association between these genes and tumors.

The deletion of chromosome 6q has been extensively mapped in a variety of tumors (11, 12). UTRN is located in this region, which encodes dystrophin. UTRN is a tumor suppressor inducing cell transformation when expressed in an antisense orientation. Studies showed decreased expression and inactivation mutations of UTRN in tumors. Expression of a wild-type UTRN in breast cancer cells inhibited tumor cell growth *in vitro* and reduced their tumor potential in nude mice (13). HYDIN is a gene whose impaired function has been linked to abnormal ciliary function, dyskinesia, and brain abnormalities (14, 15). HYDIN-derived sequences are targeted by the adaptive immunity in patients with cancer (16). Somatic mutations in HYDIN were found in breast cancer samples (17–19). MUC16, encoding a type I transmembrane mucin protein (20, 21), is frequently mutated in multiple types of human cancer (22). MUC16 was reported to modulate immune response to cancer (23–25). CA125 is a repeating peptide epitope of the mucin MUC16 (26, 27). MUC16 mutations were found to be associated with higher tumor mutation load, better survival outcomes, and immune response and cell cycle pathways in gastric adenocarcinoma (28). The CCDC178 is an 867-amino acid polypeptide and belongs to the superfamily of coiled-coil domain-containing protein. CCDC178 was reported to be mutated in hepatocellular carcinoma (29) and gastric carcinoma (30). CCDC178 associated with BRCA1-associated protein 2 (BRAP2), a negative regulator of extracellular signal-regulated kinase (ERK) pathway, and promoted its degradation (31). CCDC178 deficiency impaired the ERK activation in hepatocellular carcinoma (31). In our study, CCDC178 mutations were significantly associated with a poor survival outcome in gastric adenocarcinoma. The relative transcriptional level of CCDC178 was significantly downregulated in several types of carcinoma compared with adjacent non-cancerous tissues in TCGA cohorts (**Supplementary Figure 1**).

Compared with mutation detection, measurement of gene expression in cross-platform is unstable. Due to the lack of reproducibility and standardization, its clinical application may be limited. Recently, high-throughput sequencing technologies have been widely utilized in clinical cancer research. Compared to normal tissues, many high/low expressed proteins and mutated genes in tumor cells were identified. Combining several altered genes together may be feasible in predicting gastric adenocarcinoma risk and prognosis. In our study, the

four genes (UTRN, MUC16, CCDC178, and HYDIN) were mutated in 177 patients (40.5%) in TCGA STAD cohort. The risk score is powerful and accurate in prognostic stratification. Our work provided an advanced method toward clinical applications of gene mutation profiling in STAD, especially in future personalized prediction and precision medicine.

However, our study has several limitations. The number of samples with follow-up data in the Tianjin cohort was limited. No CCDC178 mutation was detected in the Tianjin cohort. Gastric adenocarcinoma is a heterogeneous disease. Molecular subtyping can encompass this heterogeneity and provide useful clinical information. Prognostic tool constructed on the basis of anatomic site, histopathology, and molecular subtype may be more powerful and accurate. Considering the number of samples, anatomic site, histopathology, and molecular subtype were not included in this study. So, more prospective studies are necessary to further validate the reliability and stability of this risk score.

CONCLUSIONS

A four-mutated gene risk formula was developed that correlated with the overall survival of patients with gastric adenocarcinoma using a multivariable Cox regression model. In two independent genomic datasets from TCGA and Tianjin cohorts, low risk scores were associated with higher tumor mutation loads and improved outcome in patients with gastric adenocarcinoma. This finding may have implications for prognostic prediction and therapeutic guidance for gastric adenocarcinoma.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: <https://portal.gdc.cancer.gov> <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422640112/-/DCSupplemental>.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BZ: statistical analysis, analysis and interpretation of data, drafting of the manuscript. YL: critical revision of the manuscript for important intellectual content. LY: critical revision of the manuscript for important intellectual content. YC: critical revision of the manuscript for important intellectual

content. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.584213/full#supplementary-material>

REFERENCES

- Song Z, Wu Y, Yang J, Yang D, Fang X. Progress in the Treatment of Advanced Gastric Cancer. *Tumour Biol: J Int Soc Oncodevelopmental Biol Med* (2017) 39(7):1010428317714626. doi: 10.1177/1010428317714626
- Gayed DT, Wodeyar J, Wang ZX, Wei X, Yao YY, Chen XX, et al. Prognostic Values of Inhibitory kappaB Kinases mRNA Expression in Human Gastric Cancer. *Biosci Rep* (2019) 39(1):BSR20180617. doi: 10.1042/BSR20180617
- Jin C, Lin JR, Ma L, Song Y, Shi YX, Jiang P, et al. Elevated Spondin-2 Expression Correlates With Progression and Prognosis in Gastric Cancer. *Oncotarget* (2017) 8(6):10416–24. doi: 10.18632/oncotarget.14423
- Erichsen R, Kelsh MA, Oliner KS, Nielsen KB, Froslev T, Laenkholm AV, et al. Prognostic Impact of Tumor MET Expression Among Patients With Stage IV Gastric Cancer: A Danish Cohort Study. *Ann Epidemiol* (2016) 26(7):500–3. doi: 10.1016/j.annepidem.2016.05.002
- Liu X, Xu D, Huang C, Guo Y, Wang S, Zhu C, et al. Regulatory T Cells and M2 Macrophages Present Diverse Prognostic Value in Gastric Cancer Patients With Different Clinicopathologic Characteristics and Chemotherapy Strategies. *J Trans Med* (2019) 17(1):192. doi: 10.1186/s12967-019-1929-9
- Chen K, Yang D, Li X, Sun B, Song F, Cao W, et al. Mutational Landscape of Gastric Adenocarcinoma in Chinese: Implications for Prognosis and Therapy. *Proc Natl Acad Sci USA* (2015) 112(4):1107–12. doi: 10.1073/pnas.1422640112
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118
- Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *Omics: J Integr Biol May* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
- Cancer Genome Atlas Research N. Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature* (2014) 513(7517):202–9. doi: 10.1038/nature13480
- Li X, Wu WK, Xing R, Wong SH, Liu Y, Fang X, et al. Distinct Subtypes of Gastric Cancer Defined by Molecular Characterization Include Novel Mutational Signatures With Prognostic Capability. *Cancer Res* (2016) 76(7):1724–32. doi: 10.1158/0008-5472.CAN-15-2443
- Mertens F, Johansson B, Hoglund M, Mitelman F. Chromosomal Imbalance Maps of Malignant Solid Tumors: A Cytogenetic Survey of 3185 Neoplasms. *Cancer Res* (1997) 57(13):2765–80.
- Offit K, Parsa NZ, Gaidano G, Filippa DA, Louie D, Pan D, et al. 6q Deletions Define Distinct Clinico-Pathologic Subsets of Non-Hodgkin's Lymphoma. *Blood* (1993) 82(7):2157–62. doi: 10.1182/blood.V82.7.2157.2157
- Li Y, Huang J, Zhao YL, He J, Wang W, Davies KE, et al. UTRN on Chromosome 6q24 Is Mutated in Multiple Tumors. *Oncogene* (2007) 26(42):6220–8. doi: 10.1038/sj.onc.1210432
- Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent Reciprocal 1q21.1 Deletions and Duplications Associated With Microcephaly or Macrocephaly and Developmental and Behavioral Abnormalities. *Nat Genet* (2008) 40(12):1466–71. doi: 10.1038/ng.279
- Olbrich H, Schmidts M, Werner C, Onoufriadis A, Loges NT, Raidt J, et al. Recessive HYDIN Mutations Cause Primary Ciliary Dyskinesia Without Randomization of Left-Right Body Asymmetry. *Am J Hum Genet* (2012) 91(4):672–84. doi: 10.1016/j.ajhg.2012.08.016
- Laske K, Shebzukhov YV, Grosse-Hovest L, Kuprash DV, Khlgatian SV, Koroleva EP, et al. Alternative Variants of Human HYDIN Are Novel Cancer-Associated Antigens Recognized by Adaptive Immunity. *Cancer Immunol Res* (2013) 1(3):190–200. doi: 10.1158/2326-6066.CIR-13-0079
- Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The Landscape of Cancer Genes and Mutational Processes in Breast Cancer. *Nature* (2012) 486(7403):400–4. doi: 10.1038/nature11017
- Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence Analysis of Mutations and Translocations Across Breast Cancer Subtypes. *Nature* (2012) 486(7403):405–9. doi: 10.1038/nature11154
- Zhang Y, Cai Q, Shu XO, Gao YT, Li C, Zheng W, et al. Whole-Exome Sequencing Identifies Novel Somatic Mutations in Chinese Breast Cancer Patients. *J Mol Genet Med Int J Biomed Res* (2015) 9(4):183. doi: 10.4172/1747-0862.1000183
- Xiang X, Feng M, Felder M, Connor JP, Man YG, Patankar MS, et al. HN125: A Novel Immunoadhesin Targeting MUC16 With Potential for Cancer Therapy. *J Cancer* (2011) 2:280–91. doi: 10.7150/jca.2.280
- Felder M, Kapur A, Gonzalez-Bosquet J, Horibata S, Heintz J, Albrecht R, et al. MUC16 (CA125): Tumor Biomarker to Cancer Therapy, a Work in Progress. *Mol Cancer* (2014) 13:129. doi: 10.1186/1476-4598-13-129
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational Heterogeneity in Cancer and the Search for New Cancer-Associated Genes. *Nature* (2013) 499(7457):214–8. doi: 10.1038/nature12213
- Rump A, Morikawa Y, Tanaka M, Minami S, Umesaki N, Takeuchi M, et al. Binding of Ovarian Cancer Antigen CA125/MUC16 to Mesothelin Mediates Cell Adhesion. *J Biol Chem* (2004) 279(10):9190–8. doi: 10.1074/jbc.M312372200
- Hatrup CL, Gendler SJ. Structure and Function of the Cell Surface (Tethered) Mucins. *Annu Rev Physiol* (2008) 70:431–57. doi: 10.1146/annurev.physiol.70.113006.100659
- Comamala M, Pinard M, Theriault C, Matte I, Albert A, Boivin M, et al. Downregulation of Cell Surface CA125/MUC16 Induces Epithelial-to-Mesenchymal Transition and Restores EGFR Signalling in NIH : OVCAR3 Ovarian Carcinoma Cells. *Br J Cancer* (2011) 104(6):989–99. doi: 10.1038/bjc.2011.34
- O'Brien TJ, Beard JB, Underwood LJ, Dennis RA, Santin AD, York L. The CA 125 Gene: An Extracellular Superstructure Dominated by Repeat Sequences. *Tumour Biol: J Int Soc Oncodevelopmental Biol Med* (2001) 22(6):348–66. doi: 10.1159/000050638
- Yin BW, Lloyd KO. Molecular Cloning of the CA125 Ovarian Cancer Antigen: Identification as a New Mucin, MUC16. *J Biol Chem* (2001) 276(29):27371–5. doi: 10.1074/jbc.M103554200
- Li X, Pasche B, Zhang W, Chen K. Association of MUC16 Mutation With Tumor Mutation Load and Outcomes in Patients With Gastric Cancer. *JAMA Oncol* (2018) 4(12):1691–8. doi: 10.1001/jamaoncol.2018.2805
- Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, et al. Exome Sequencing of Hepatitis B Virus-Associated Hepatocellular Carcinoma. *Nat Genet* (2012) 44(10):1117–21. doi: 10.1038/ng.2391
- Zhang J, Huang JY, Chen YN, Yuan F, Zhang H, Yan FH, et al. Whole Genome and Transcriptome Sequencing of Matched Primary and Peritoneal Metastatic Gastric Carcinoma. *Sci Rep* (2015) 5:13750. doi: 10.1038/srep13750
- Hu X, Zhao Y, Wei L, Zhu B, Song D, Wang J, et al. CCDC178 Promotes Hepatocellular Carcinoma Metastasis Through Modulation of Anoikis. *Oncogene* (2017) 36(28):4047–59. doi: 10.1038/onc.2017.10

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Li, Yang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License

(CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.