Analyzing TCGA Data to Identify Gene Mutations Linked to Hepatocellular Carcinoma in Asians

Tane Kim\textsuperscript{a}  Danny Issa\textsuperscript{b}  Mykola Onyshchenko\textsuperscript{c}

\textsuperscript{a}Orange County School of the Arts, Santa Ana, CA, USA; \textsuperscript{b}Division of Digestive Diseases, Department of Medicine, UCLA School of Medicine, Los Angeles, CA, USA; \textsuperscript{c}Hematology-Oncology Division, Department of Medicine, Harbor-UCLA/Lundquist Institute, Torrance, CA, USA

Keywords
Asian · Hepatocellular carcinoma · The Cancer Genome Atlas · Bioinformatics · Prognostic marker

Abstract

Introduction: Liver cancer is the sixth most common and second most fatal type of cancer worldwide. Few treatment options are available as patients with liver cancer are often diagnosed in an advanced stage due to a lack of clinical symptoms. Effectively preventing and treating liver cancer relies heavily on early diagnosis; early diagnosis results from identifying and monitoring high-risk patients. Epigenetic risk factors, such as hepatitis B, hepatitis C, cirrhosis, nonalcoholic fatty liver disease, and alcohol/tobacco abuse, are highly prevalent in Asia and likely cause Asians to have a higher incidence and mortality rate of liver cancer. While these acquired risk factors are relatively well understood, the underlying genetic background of liver cancer in Asians has not been well established or correlated with clinical outcomes.

Methods: In this study, we accessed The Cancer Genome Atlas (TCGA) hepatocellular carcinoma clinical and mutation data through TCGAbiolinksGUI.

Results: We found that mutations in five genes (\textit{TP53}, \textit{TTN}, \textit{OBSCN}, \textit{MUC5B}, \textit{CSMD1}) were statistically linked with increased mortality in Asians compared to non-Asians, four of which (\textit{TTN}, \textit{OBSCN}, \textit{MUC5B}, \textit{CSMD1}) were also more prevalent in the Asian population. Within the Asian cohort, two gene mutations (\textit{TTN}, \textit{HMCN1}) were statistically linked with worse outcomes. We also found that the \textit{TP53} mutation predicts worse outcomes within the non-Asian cohort but not within the Asian cohort.

Discussion/Conclusion: Our findings can improve cancer care in the Asian population through better disease prognostication, evaluations for potential targeted therapy, and a deeper understanding of liver cancer pathogenesis.

Introduction

Liver cancer is the sixth most common and second most fatal type of cancer [1]. From 1990 to 2015, the incidence rate of liver cancer has increased by 75% worldwide [2]. The overall 5-year relative survival rate, based on the Surveillance, Epidemiology, and End Results Program, is 34% for localized disease, 12% for locally advanced or with lymph node involvement, and 3% for distant metastases [3]. The average 5-year survival rate of liver cancer is 19%, as determined by the American Cancer Society using data from 2010 to 2016 [3, 4].

The most common types of liver cancer are hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). These two subtypes make up 90% and 10% of liver cancer, respectively [5]. The primary acquired risk factors
factors for developing HCC and ICC described in the literature are hepatitis B, hepatitis C, parasite infection, nonalcoholic fatty-liver disease, liver cirrhosis, and alcohol and tobacco use [6, 7]. An exposure to hepatic carcinogens (e.g., aflatoxins, arsenic) is also strongly correlated with an onset of liver cancer [8].

A cancer incidence report published by Ferlay et al.[1] in 2015 concluded that liver cancer is significantly more prevalent in less-developed regions, such as Southeast Asia and Africa [1, 9]. These regions have significantly higher reported cases of hepatitis B, hepatitis C [10], liver fluke infection [5], and aflatoxins exposure [11] than developed countries. Recently NAFDL prevalence in Asia has approached similar levels to those in Western countries and was reported higher in certain areas and obese children [12, 13]. Even though these risk factors are endemic to Asia and Africa, liver cancer is nearly twice as common and fatal in Asian Americans than non-Asian Americans [14, 15]. This phenomenon suggests that regardless of exposure to environmental risk factors, Asians are likely still highly susceptible to liver cancer, possibly due to genetic background. This study mines The Cancer Genome Atlas (TCGA) for gene mutations linked with worse outcomes in Asian patients with HCC.

Materials and Methods

Data Download and Analysis
TCGA is a publicly available database that provides cancer genomics data to researchers. Docker Desktop was launched in order to provide a local server for the RStudio virtual environment. Powershell was used to run the RStudio local server, which was in turn used to run TCGAbioliinksGUI. Patient mutation and clinical data from TCGA were downloaded through TCGAbioliinksGUI on July 3, 2020. These data were uploaded to the GenePattern Jupyter Notebook server and analyzed using the pandas, NumPy, Matplotlib, seaborn, and SciPy libraries in Python.

SciPy, a Python library used for scientific computing, was used to analyze the prognostic significance of these 24 common genes on survival using a two-tailed test. KaplanMeierFitter, a program that utilizes the Kaplan-Meier method, was used to graph the 5-year survival rates.

Selection of Patient Sample
The number of HCC patients registered in TCGA is 375, most of whom are White and Asian (187 and 160, respectively) (shown in Table 1). The overall survival rates of Asian and non-Asian patients were compared; at around 750 days since diagnosis, the Asian survival curve began to flatten noticeably. Furthermore, the estimated 5-year survival rates of Asian and non-Asian patients were 2–3 times higher than published statistics [4]. These discrepancies suggest poor follow-up with living patients (shown in Fig. 1).

When only the survival rates of deceased patients were compared, the resulting survival graphs better represented known trends (shown in Fig. 2) [4]. Accordingly, this study only analyzed deceased Asian and non-Asian HCC patients.

Survival Analysis of Asian and Non-Asian
Seaborn was used to analyze the survival rate differences between all Asian and non-Asian HCC patients.

Survival Analysis of Deceased Asian and Non-Asian Patients
Kaplan-Meier and Matplotlib were used to investigate the survival rate differences between deceased Asian and non-Asian HCC patients.

Gene Mutation Effects on the Survival Rate in Deceased Asian and Non-Asian Patients
The 24 most common mutated genes in deceased Asian HCC patients were identified. Kaplan-Meier and Matplotlib were used to compare the survival rates of Asian and non-Asian patients with these genes.

Gene Mutation Effects on the Survival Rate in Deceased Asian Patients
The 24 genes were used to determine their effect on Asian prognosis. Kaplan-Meier and Matplotlib were used to compare the survival rate of Asian patients with these genes and Asian patients without these genes.

Statistical Analysis
SciPy was used to conduct Welch’s t test (unequal variances t test), with p values <0.05 considered to be statistically significant.

### Table 1. HCC patient demographic information

| Race                         | Gender | Vital status | Alive | Dead | Total |
|------------------------------|--------|--------------|-------|------|-------|
| American Indian or Alaska Native | Male   | 2            | 0     | 0    | 2     |
|                              | Female | 0            | 0     | 0    | 0     |
| Asian                        | Male   | 91           | 35    | 126  |
|                              | Female | 25           | 9     | 34   |
| Black or African American    | Male   | 8            | 5     | 13   |
|                              | Female | 2            | 1     | 3    |
| Not reported                 | Male   | 4            | 3     | 7    |
|                              | Female | 1            | 2     | 3    |
| White                        | Male   | 67           | 38    | 105  |
|                              | Female | 43           | 39    | 82   |
| Total                        |        | 243          | 132   | 375  |
Fig. 1. The overall survival rate of Asian versus non-Asian HCC patients.

Fig. 2. The overall survival rate of deceased Asian versus deceased non-Asian HCC patients.
Results

Survival Analysis of Deceased Asian and Non-Asian Patients

The clinical data showed that deceased Asian HCC patients had a significantly lower survival rate than deceased non-Asian HCC patients (shown in Table 2, shown in Fig. 2, \( p < 0.000001 \)).

Gene Mutation Effects on the Survival Rate in Deceased Asian and Non-Asian Patients

The 24 most frequently mutated genes in deceased Asian HCC patients were identified and graphed by prevalence. The prevalence of this mutation in non-Asians was graphed for comparison. The majority of mutations had a higher prevalence among Asian patients compared to non-Asian patients (shown in Fig. 3).

Single-nucleotide polymorphism mutations in \( TP53 \), \( TTN \), \( OBSCN \), \( MUC5B \), and \( CSMD1 \) genes were statistically associated with worse overall survival in Asian HCC patients than non-Asian HCC patients, with a \( p \) value <0.05 considered to be statistically significant. Hemicentin 1 (\( HMCN1 \)), a significant gene mutation within the Asian population, had a trend toward worse survival than non-Asians, but the difference was not statistically significant (shown in Fig. 4).

Gene Mutation Effects on the Survival Rate in Deceased Non-Asian Patients

We further studied if identified mutations in \( TP53 \), \( TTN \), \( OBSCN \), \( MUC5B \), \( CSMD1 \), and \( HMCN1 \) predict...
Fig. 4. Survival rates of Asian versus non-Asian HCC patients with mutations of TP53 (a) TTN (b) OBSCN (c) MUC5B (d) CSMD1 (e) HMCN1 (f).

(Figure continued on next pages.)
Gene Mutations Linked to Hepatocellular Carcinoma in Asians

CSMD1 | Survival Rate Asian vs Non-Asian | p-value = 0.001313

HMCN1 | Survival Rate Asian vs Non-Asian | p-value = 0.1509
Fig. 5. Survival rates of non-Asian HCC patients with and without mutations of TP53 (a) TTN (b) OBSCN (c) MUC5B (d) CSMD1 (e) HMCN1 (f).

(Figure continued on next pages.)
overall survival within Asian and non-Asian populations separately. The results in Figure 5 show that mutations in TTN, OBSCN, MUC5B, CSMD1, and HMCN1 genes were not significant in the non-Asian sample group. In contrast, the mutation in the TP53 gene was statistically significantly associated with worse mortality among non-Asians.

We then studied if the identified six gene mutations were associated with a worse outcome within the Asian patient cohort. Mutations in TTN led to significantly lower survival rates among Asian patients (shown in Fig. 6a). Another gene mutation, HMCN1, was also linked to a significantly lower survival rate within the Asian HCC patient cohort (shown in Fig. 6b).

**Statistical Analysis**

All calculated *p* values < 0.05 were considered statistically significant in this study. Table 3 shows a summary of the results.

### Table 3. Significant genes in HCC survival based on TCGA data

| Group 1          | Sample size | Days to death, mean | Group 2          | Sample size | Days to death, mean | *p* value |
|------------------|-------------|---------------------|------------------|-------------|---------------------|-----------|
| Asian versus non-Asian |             |                     |                  |             |                     |           |
| Asian            | 44          | 325                 | Non-Asian        | 83          | 866                 | <0.00001  |

**Mutated gene comparison**

| Gene          | Sample size | Days to death, mean | *p* value |
|---------------|-------------|---------------------|-----------|
| Asian with TP53 | 18          | 270                 | 0.029     |
| Asian with TTN  | 16          | 199                 | 0.002     |
| Asian with OBSCN | 7           | 195                 | 0.017     |
| Asian with MUC5B | 6           | 242                 | 0.012     |
| Asian with CSMD1 | 6           | 194                 | 0.001     |
| Asian with HMCN1 | 6           | 107                 | 0.151     |

**Mutated gene comparison among non-Asians**

| Gene          | Sample size | Days to death, mean | *p* value |
|---------------|-------------|---------------------|-----------|
| Non-Asian with TP53 | 21          | 570                 | 0.18      |
| Non-Asian with TTN   | 22          | 840                 | 0.86      |
| Non-Asian with OBSCN | 12          | 732                 | 0.467     |
| Non-Asian with MUC5B | 3           | 732                 | 0.296     |
| Non-Asian with CSMD1 | 7           | 1,043               | 0.320     |
| Non-Asian with HMCN1 | 5           | 456                 | 0.089     |

**Mutated gene comparison among Asians**

| Gene          | Sample size | Days to death, mean | *p* value |
|---------------|-------------|---------------------|-----------|
| Asian with TP53 | 18          | 271                 | 0.301     |
| Asian with TTN  | 16          | 199                 | 0.019     |
| Asian with OBSCN | 7           | 195                 | 0.065     |
| Asian with MUC5B | 6           | 243                 | 0.239     |
| Asian with CSMD1 | 6           | 195                 | 0.074     |
| Asian with HMCN1 | 6           | 108                 | 0.003     |

### Discussion

This study hypothesizes that lower survivability in Asian HCC patients than non-Asian HCC patients may be linked to somatic gene mutation signatures rather than epigenetic differences.

To begin, the initial clinical and mutation data needed to be filtered. Only HCC patient data were analyzed in this study because the sample size of ICC patients was too small to produce significant results. We assumed that the results are still relevant as HCC accounts for 90% of liver cancer [16]. Furthermore, the survival rates of HCC patients were significantly higher than published literature; this indicates that there may have been little follow-up with living patients, with a large majority of patients incorrectly marked as alive in the database many years later. This study only analyzed deceased HCC patient data in order to prevent skewed survival curves.

In the patient group of deceased Asian and non-Asian HCC patients, the Asian patients’ days to death since diagnosis (mean = 325) is statistically shorter (*p* value...
Fig. 6. Survival rates of Asian HCC patients with and without mutations of TP53 (a) TTN (b) OBSCN (c) MUC5B (d) CSMD1 (e) HMCN1 (f).

(Figure continued on next pages.)
Gene Mutations Linked to Hepatocellular Carcinoma in Asians

Gastrointest Tumors 2022;9:43–58
DOI: 10.1159/000524576
Gene Mutations Linked to Hepatocellular Carcinoma in Asians

<0.000001) than non-Asian patients (mean = 866). Epigenetic factors, such as diet, rates of chronic liver diseases, and access to healthcare may play a significant role in the unfavorable prognosis for Asian patients. However, distinct somatic mutational profiles may contribute to the outcomes as well.

After identifying the 24 most frequently mutated genes in deceased Asian HCC patients, the survival rates of deceased Asian HCC patients with these mutations were compared with those of deceased non-Asian HCC patients with these gene mutations. Five genes were statistically linked to worse survival in Asians than non-Asians. These five genes are TP53, TTN, OBSCN, MUC5B, and CSMD1; HMCN1 had a trend toward worse survival. Also, all identified gene mutations except OBSCN had a higher prevalence among Asians compared to non-Asians.

To see whether the identified genes independently determined the outcome, rather than epigenetic race-related factors, we studied the association between identified gene mutations and outcomes within Asian and non-Asian patient cohorts separately. In the non-Asian cohort, among six genes, only the TP53 mutation was linked to worse survival; TP53 mutation, in general, is well known poor prognostic marker in a variety of tumor histologies. Interestingly, when we analyzed these six genes in the Asian cohort, TP53 mutation did not have a prognostic value. However, two genes were statistically significantly linked to worse outcomes in Asian HCC patients versus Asians without these mutations. These two genes mutations are TTN (p = 0.019) and HMCN1 (p = 0.003).

Among all analyzed gene mutations, TTN was the only mutated gene shown to be a statistically significant negative prognostic marker when comparing Asians versus non-Asians and within the Asian patient cohort separately. TTN encodes for Titin, a large protein found in cardiac and skeletal muscles; TTN mutations are known to cause cardiomyopathy and muscular dystrophy [17].

TTN mutations have also been correlated with increased tumor mutation burden in cancer patients with the mutation. Melanoma and lung cancer patients with TTN mutations have shown better responses to immunotherapies with more prolonged overall and progression-free survival. However, solid tumor patients with this mutation who did not receive immunotherapy showed poor overall and disease-free survival [18].

HMCN1 encodes for immunoglobulin in adhesive and flexible cell junctions; however, its detailed function remains unknown. Mutations in HMCN1 have been linked to cancer cell invasion and metastasis, and high mutation levels have been found in patients with head and neck squamous cell carcinoma [19].

Multiple studies, such as the ones published by Rao et al. [8] and Li et al. [20], indicates that TP53 is a commonly mutated gene linked to HCC and can be used as a biomarker for prevention, prognosis, and treatment. This study identifies new gene mutations that can serve as prognostic biomarkers for Asian HCC patients. In particular, TTN and HMCN1’s high prevalence in Asian patients (36.4% and 13.6%, respectively) make them critical gene mutations with a potential prognostication that may correlate with outcomes of various HCC treatment regimens. More research is needed to understand TTN mutation’s role in HCC pathogenesis and explain the mechanism of its negative association among Asian patients.

Our results are limited to the data available in TCGA. The small size of the database, outliers, limited survival data, unknown potentially confounding factors, and the lack of epigenetic data impact the analysis and limit our conclusions. We may not know the causality links between epigenetic factors of HCC carcinogenesis and identified somatic mutation evolution. Nevertheless, TCGA remains the largest validated cancer genome database supervised by National Cancer Institute and widely accepted among cancer researchers. We found that mutations in five genes (TP53, TTN, OBSCN, MUC5B, CSMD1) were statistically linked with increased mortality in Asians compared to non-Asians, four of which (TTN, OBSCN, MUC5B, CSMD1) were also more prevalent in the Asian population. Performing the analysis in the Asian patient cohort separately, mutations in TTN and HMCN1 genes appear to be independent negative prognostic markers for Asians. Other gene mutations, such as OBSCN and CSMD1 that showed a trend toward worse survival, may be important, but the number of patients in the TCGA dataset limits this conclusion. Future studies using multiple cancer databases could help confirm the validity and applicability of the findings in this study.

**Conclusion**

In summary, a list of the 24 most frequently mutated genes found in deceased Asian HCC patients was extracted from TCGA. Mutations in five of these genes (TP53, TTN, OBSCN, MUC5B, and CSMD1) are correlated with shorter life expectancy in Asian HCC patients compared to non-Asian HCC patients. Mutations in four of these genes (TTN, OBSCN, MUC5B, and CSMD1), as well as
*HMCDN1* are only significant in Asian HCC patients. Furthermore, Asian HCC patients with *TTN* and *HMCDN1* mutations have shorter life expectancy than Asian HCC patients without these mutations. This study identified multiple genetic biomarkers that can aid in the recognition, surveillance, prognosis, and gene therapy of HCC. Further research is needed to understand the relationship between these gene mutations and epigenetic risk factors in HCC pathogenesis.

**Statement of Ethics**

This study uses data from public-use data sets and does not require IRB review.

**Conflict of Interest Statement**

The authors have no conflicts of interest to declare.

**References**

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015 Mar 1;136(5):E359–386.
2. American Cancer Society. Cancer facts & statistics. 2021 [cited 2021 Jul 14]. Available from: https://cancerstatisticscenter.cancer.org/module/yg6E0ZLc.
3. American Cancer Society. Liver cancer early detection, diagnosis, and staging. 2021 Jan 29. Available from: https://www.cancer.org/cancer/liver-cancer/detection-diagnosis-staging/early-detection.html.
4. American Cancer Society. Cancer facts & statistics. 2020 [cited 2021 Jul 14]. Available from: https://cancerstatisticscenter.cancer.org/module/CNgHqCms.
5. Alsaleh M, Leftley Z, Barbera TA, Sithithaworn P. Non-alcoholic fatty liver disease (NAFLD) and fibrosis: the role of liver biopsy. *World J Gastroenterol*. 2012 Dec 14;18(45):6997–7004. Available from: https://www.wjgnet.com/wjg8223-7700/fullpaper/WJ82238464.pdf.
6. Mohamad B, Shah V, Onyshchenko M, Elshamy M, Aucejo F, Lopez R, et al. Characterization of hepatocellular carcinoma (HCC) in non-alcoholic fatty liver disease (NAFLD): patients without cirrhosis. *Hepatol Int*. 2016 Jul;10(4):632–9.
7. American Cancer Society. Liver cancer risk factors. 2019 [cited 2021 Jul 27]. Available from: https://www.cancer.org/cancer/liver-cancer/causes-risks-prevention/risk-factors.html.
8. Rao CV, Asch AS, Yamada HY. Frequently mutated genes/pathways and genomic instability as prevention targets in liver cancer. *Carcinogenesis*. 2017 Jan;38(1):2–11.
9. Petrick JL, Braunlin M, Laversonne M, Valery PC, Bray F, McGlynn KA. International trends in liver cancer incidence, overall and by histologic subtype, 1978–2007. *Int J Cancer*. 2016 Oct 1;139(7):1534–45.
10. Jefferies M, Rauff B, Rashid H, Lam T, Rafiq S. Update on global epidemiology of viral hepatitis and preventive strategies. *World J Clin Cases*. 2018 Nov 6 [cited 2021 Jul 25]; 6(13):589–99. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6304240/.
11. Paliyaguru DL, Wu F. Global geographical overlap of aflatoxin and hepatitis C: controlling risk factors for liver cancer worldwide. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*. 2013;30(3):534–40.
12. Li J, Zou B, Yeo YH, Feng Y, Xie X, Lee DH, et al. Prevalence, incidence, and outcome of non-alcoholic fatty liver disease in Asia, 1999–2019: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2019 May;4(5):389–98.
13. Fan JG, Kim SU, Wong VW. New trends on obesity and NAFLD in Asia. *J Hepatol*. 2017 Oct;67(4):862–73.
14. American Cancer Society. Key statistics about liver cancer2021 [cited 2021 Jul 22]. Available from: https://www.cancer.org/cancer/liver-cancer/about/what-is-key-statistics.html.
15. The Office of Minority Health. Chronic liver disease and Asian Americans. 2020 [cited 2021 Jul 27]. Available from: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=47.
16. American Cancer Society. What is liver cancer? 2019 [cited 2021 Jul 14]. Available from: https://www.cancer.org/cancer/liver-cancer/about/what-is-liver-cancer.html.
17. MedlinePlus. TTN gene: MedlinePlus genetics. 2021 [cited 2021 Jul 28]. Available from: https://medlineplus.gov/genetics/gene/tnn/.
18. Jia Q, Wang J, He N, He J, Zhu B. Titin mutation associated with responsiveness to checkpoint blockades in solid tumors. *JCI Insight*. 2019 May 16 [cited 2021 Jul 28];4(10):e127901. Available from: https://doi.org/10.1172/jci.insight.127901.
19. Kikutake C, Yoshihara M, Sato T, Saitsu D, Suyama M. Intratumor heterogeneity of HMCN1 mutant alleles associated with poor prognosis in patients with breast cancer. *Onco Targets Ther*. 2018 Sep 7 [cited 2021 Jul 28];9(70):3333–47. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6210430/.
20. Li X, Xu W, Kang W, Wong SH, Wang M, Zhou Y, et al. Genomic analysis of liver cancer unveils novel driver genes and distinct prognostic features. *Theranostics*. 2018;8(6):1740–51.

**Funding Sources**

No funding was received.

**Author Contributions**

Tane Kim conceived the project, designed the computational framework, and analyzed the data. Mykola Onyshchenko was in charge of overall planning and direction. Tane Kim and Mykola Onyshchenko wrote the manuscript with input from Danny Issa.

**Data Availability Statement**

This study is based upon data generated by the TCGA Research Network, which can be found at https://www.cancer.gov/tcga. All data generated or analyzed during this study can be found in figshare at https://doi.org/10.6084/m9.figshare.c.5666116.v1.