Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/92/20241737

Pan-cancer analysis of TP53 mutations and their association with liver cancer and aristolochic acids

Zhetao Zhao

University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom

zz6g20@soton.ac.uk

Abstract. Liver cancer is a major global health concern, as its incidence continues to rise in numerous countries, making it a leading cause of cancer-related deaths. The TP53 gene, which encodes the p53 tumor suppressor protein, plays a crucial role in regulating cell growth and division. Mutations in TP53 have been implicated in increased risk for liver cancer development. In this study, we aim to conduct a pan-cancer analysis to further investigate the potential role of TP53 mutations in various cancer types and their relationship with liver cancer. Additionally, we will explore the interplay between TP53 gene mutations, liver cancer, and exposure to Aristolochic acids (AAs). To achieve these goals, we will utilize data from The Cancer Genome Atlas (TCGA) project and the Gene Expression Omnibus (GEO) databases, enabling a comprehensive analysis of TP53 in the context of multiple cancer types.

Keywords: Liver cancer, TP53 mutations, pan-cancer analysis, Aristolochic acids, and The Cancer Genome Atlas (TCGA).

1. Introduction

This study used various methods to analyze the expression of the TP53 gene and its relationship with Aristolochic acids (AAs) and the development of liver cancer. Timer2.0's Gene_DE module and the GEPIA2 software to compare the expression of TP53 between tumor and normal tissue in various tumors from the TCGA project [1]. UALCAN portal and the CPTAC database to explore protein expression levels in various types of cancer. GEPIA2 software to obtain survival data and to analyze the relationship between TP53 expression and prognosis in patients with different types of cancer.

Gene TP53 expression was significantly higher in liver cancer and certain other types of cancer compared to normal tissue, and that high expression of TP53 was associated with a poorer prognosis in these cancers [2]. AAs were associated with an increased risk of liver cancer and TP53 mutations.

2. Methods and Concepts

2.1. Gene Expression Analysis

To begin, the expression of TP53 was compared between tumour and normal tissue in various tumours or sub-types of the TCGA project [3]. This study used Timer2.0's (Tumour immune estimation resource, version 2) Gene_DE module, which allows users to compare the expression of any gene of interest in

DOI: 10.54254/2755-2721/92/20241737

all TCGA tumours between tumour and normal tissue. The distribution of gene expression levels is visualized using box plots.

In addition, we examine the integration of the "Expression Analysis-Box Plots" module within the GEPIA2 (Gene expression profiling interactive analysis, version 2) software to generate box plots illustrating the variations in expression levels between specific tumors lacking normal tissues or with severely restricted normal tissues, and the matching normal tissues from the GTEx (Genotype-tissue expression) database. For instance, in situations of TP53, the LGG.Tumor (Low-grade gliomas) and SARC.Tumor (Sarcoma) will be applied to the GEPIA2. By configuring this module, we initially establish log2(TPM+1) as the log-scale and 0.4 as the Jitter size to align with typical data. Next, we selected a log2FC threshold of 1 and a p-value cutoff of 0.01. In order to observe the variations in gene expression between normal and cancerous tissues, we may refer to the TCGA+GTEx dataset.

Furthermore, the CPTAC (Clinical Proteomics Tumor Analysis Consortium) database via the UALCAN portal was utilized to investigate the overall protein content at the expression level of this gene. This analysis provided additional evidence supporting the high expression of TP53 in certain cancer types, distinguishing between normal tissues and primary tumors. The website UALCAN offers comprehensive tools for the investigation of protein expression by utilizing data provided from The Cancer Proteome Atlas (CPTAC) and the International Cancer Proteome Consortium (ICPC) databases. The interface enables users to investigate protein expression levels in different cancer types, such as colorectal cancer and breast cancer. This knowledge is crucial for investigating the protein's involvement in the progression of certain specific forms of cancer.

Lastly, we will once again concentrate on GEPIA2 to generate the pathological stage plot. The differential gene expression analysis method applied is one-way ANOVA, where the diseased stage is used as a variable to compute differential expression. This study aims to analyze the violin plots of TP53 expression at several clinical stages, including stage I, stage II, stage III, and stage IV.

2.2. Survival Analysis

In this analysis, we could determine the high- and low-expression cases by GEPIA2's survival map and obtain the relationship between the prognosis of patients with a different type of tumor and TP53 expression. Through GEPIA2 we could also conclude the OS(overall survival) and DFS(Disease-free survival) data for TP53 [4].

2.3. Aristolochic acids (AAs)

Aristolochic acids (AAs) are a group of toxic compounds found in plants belonging to the Aristolochiaceae family that have been linked to the development of liver cancer. In the discussion session, this paper will summarize the current evidence on the relationship between AAs and liver cancer, including the mechanisms of action, epidemiological data, and potential preventive measures.

3. Results and Data

3.1. TCGA Datasets

In Figure 1, it can be observed that the expression levels of TP53 across 33 cancer types using data from the TCGA project. Our analysis reveals that TP53 expression levels (log2 TPM) in tumor tissues are significantly higher than in corresponding control cells for various cancer types, for example Cervical Squamous Cell Carcinoma. The statistical significance is indicated by the number of stars (*, **, or ***), based on the p-value calculated using the Wilcoxon test. Columns are presented in gray when normal data is available for comparison.

DOI: 10.54254/2755-2721/92/20241737

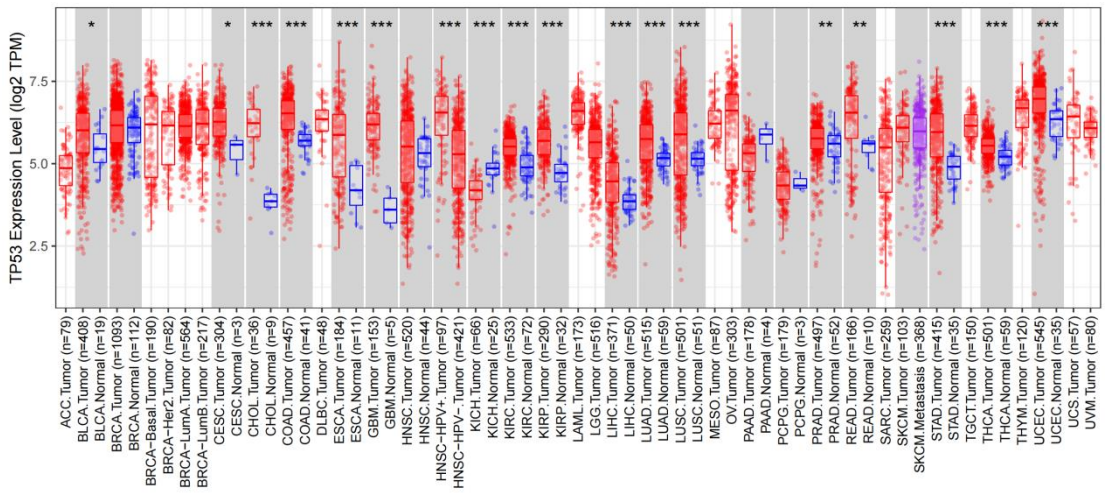

Figure 1. TP53 gene shows the expression status in 33 different cancers analyse by Timer2

3.2. TCGA + GTEx Datasets

The application of the module of the GEPIA2 has been used to insert the GTEx database, and then we can clearly analyze the different expressions of TP53 to compare with tumor tissues of varies of model, for example THYM (tumors) and UCS (uterine carcinoma).

3.3. CPTAC Datasets

For the input gene TP53, we focus on total protein to analyse the whole system. According to the findings, the primary tissues of Hepatocellular carcinoma from the CPTAC dataset had higher expression of the gene total-protein. (See Figure 2).

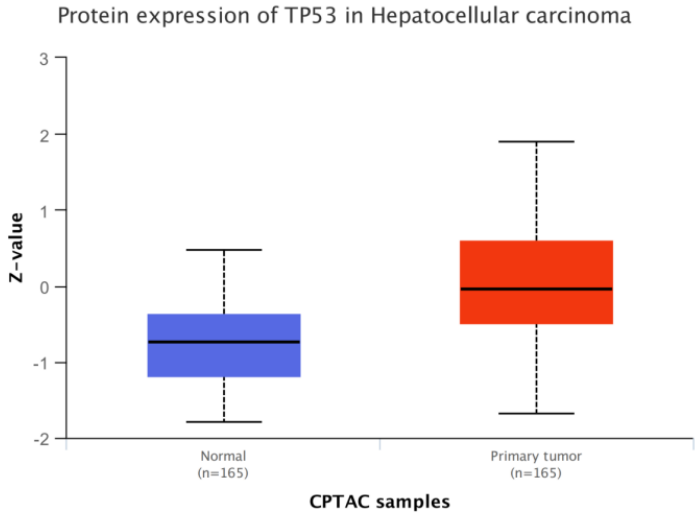

Figure 2. Total-Protein expression of TP53 in Hepatocellular carcinoma

3.4. Pathological Stage Plot

Reconsider GEPIA2. Analysis of the TP53 gene expression levels in LIHC (liver hepatocellular carcinoma) was conducted using data from the TCGA dataset. The gene expression levels were then categorized into the primary pathological phases (I, II, III, and IV). To convert the data to a log scale, the Log2 (TPM + 1) approach was employed (Refer to Figure 3).

DOI: 10.54254/2755-2721/92/20241737

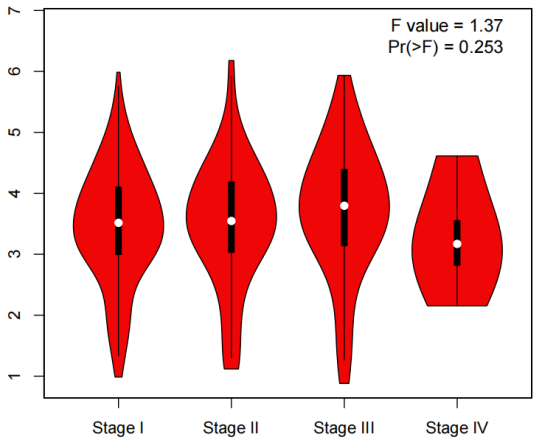

Figure 3. The expression levels of the TP53 in LIHC

3.5. Survival Data Analysis

By using the survival modulus of GEPIA2, we separate the analysis into two groups - high-expression and low-expression. The high- and low-expression cohorts were divided using cutoff-high (50%) and cutoff-low (50%) values as the expression thresholds. So that we could look at gene TP53 and the prognosis of patients with different tumors based on their expression. The outcome of the survival modulus by focusing on OS (overall survival) via the condition of highly expressed TP53 with the minimum prognosis of overall survival for all cancer types in the TCGA database. Then use the survival analysis modulus to obtain survival plots of LIHC (liver hepatocellular carcinoma) (See Figures 4). Depicts the outcome of a survival modulus with a focus on DFS (Disease-free survival) via the condition of highly expressed TP53 with the lowest DFS prognosis for all cancer types in the TCGA database. Same as previous procedures, we run survival analysis for, LIHC (Liver hepatocellular carcinoma), but this time focus on disease-free survival.

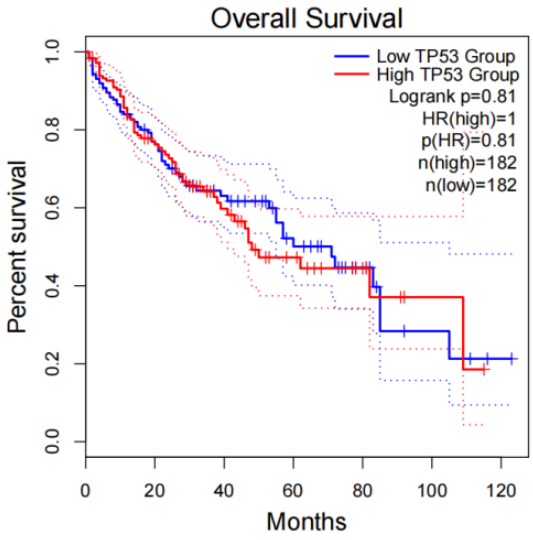

Figure 4. TP53's survival analysis of OS for LIHC (Liver hepatocellular carcinoma)

DOI: 10.54254/2755-2721/92/20241737

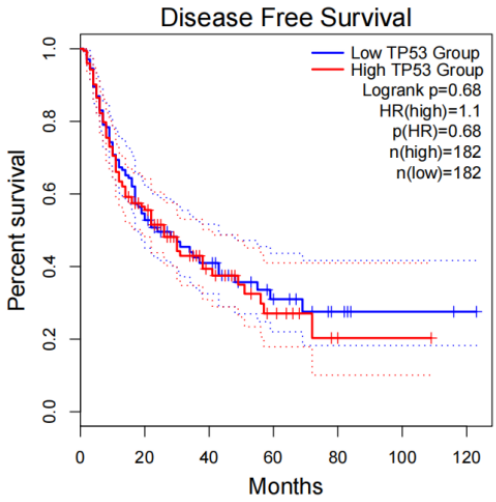

Figure 5. TP53's survival analysis of DFS for LIHC (Liver hepatocellular carcinoma)

4. Discussion

In a pan-cancer analysis, examining patterns across various cancer types provides valuable insights into cancer biology and potential treatments [5]. A key focus is the TP53 gene, which encodes the tumor suppressor protein p53, critical in cell cycle regulation and cancer prevention. Using data from TCGA, GEO, GTEx, and CPTAC, TP53 was found to be highly expressed in 33 cancers, including liver, breast, and ovarian cancers [6]. The pan-cancer analysis of TP53 reveals significant associations between TP53 mutations and clinical or molecular cancer characteristics, aiding in the identification of new therapeutic targets. Both gene expression and survival prognosis analyses contribute to the robustness of these findings, warranting further research.

Liver cancer in China, characterized by high incidence and low survival rates, is partly linked to the use of traditional medicines containing Aristolochic acids (AAs). These toxic compounds, found in plants from the Aristolochiaceae family, have been associated with TP53 gene mutations, increasing cancer risk. Additionally, AAs contribute to liver and kidney damage. Given these risks, government regulation to restrict AAs in traditional medicine is essential to protect public health and reduce cancer incidence related to TP53 mutations.

5. Conclusion

In conclusion, this research focused on exploring the potential role of TP53 in liver cancer and other types of cancer. We discovered that TP53 expression was much higher in liver cancer and some other forms of cancer relative to normal tissue, and that high TP53 expression was linked to a worse prognosis in these malignancies. Additionally, Aristolochic acids (AAs) were found to be associated with an increased risk of liver cancer and TP53 mutations. This research highlights the importance of TP53 in cancer development and provides valuable insights into potential prevention and treatment strategies for liver cancer and other types of cancer.

References

[1] K. Tomczak, P. Czerwińska, M. Wiznerowicz The cancer genome atlas (TCGA): an immeasurable source of knowledge Contemp. Oncol. (Pozn), 19 (2015), pp. A68-A77

[2] Poon, Song Ling, et al. "Genome-Wide Mutational Signatures of Aristolochic Acid and Its Application as a Screening Tool." Science Translational Medicine, vol. 5, no. 197, 7 Aug. 2013.

[3] Soussi, T, and K G Wiman. "TP53: An Oncogene in Disguise." Cell Death & Differentiation, vol. 22, no. 8, 29 May 2015, pp. 1239–1249.

DOI: 10.54254/2755-2721/92/20241737

[4] E. Clough, T. Barrett The gene expression omnibus database Methods Mol. Biol., 1418 (2016), pp. 93-110
[5] Zhang, J., et al. “International Cancer Genome Consortium Data Portal--a One-Stop Shop for Cancer Genomics Data.” Database, vol. 2011, no. 0, 18 Sept. 2011, pp. bar026–bar026.
[6] A. Blum, P. Wang, J.C. Zenklusen SnapShot: TCGA-analyzed tumors Cell., 173 (2018), p. 530