TP53 and Survival of Young Breast Cancer Patients

Introduction

According to the CDC, 264,121 new breast cancer cases for females are reported in the US in 2019, while the annual rate of new cases has not decreased much in the past twenty years and the total number of new cases and deaths has been increasing steadily with the increasing population. Now, breast cancer is becoming a major cause of death for women aging between 40 to 55 years, and single women or women who gave birth to few children have a higher risk of developing breast cancer (Akram et al, 2017). Another research suggests that 45 should be the cut-off age for breast cancer because the disease would demonstrate distinct clinicopathologic qualities (Nie et al, 2017). This report aims to find the gene with the highest mutation rate in the younger breast cancer patient population and the influence of the mutation on survival rate. The data used is from TCGA, a program that provides a large number of samples of multi-omic data on patients with a variety of cancers. This report will focus on the genomics and transcriptomics of breast cancer to gain a more comprehensive understanding of the disease. Specifically, various methods will be used to examine the association between the most mutated gene, survival status of patients, and the age category of patients. The goal is to discover the difference in breast cancer diagnosed in patients from different age groups, so future analysis can focus on deriving different treatments specific to each age group.

Methods

First, a mask will be created using the original clinical data to filter out patients that are below 45 years old. Next, query the MAF files from the TCGA website using “TCGA-BRCA”, create a MAF object using the maftools library, and the same filtering method would be applied

to create a subset from the MAF object with the genomic data using the clinical data dataframe in the MAF object. From the oncoplot of the most mutated genes in the subset, I found TP53 to be the one with the highest mutation rate (Fig 1). Then, using survival and survminer packages, a maf survival analysis was performed and a Kaplan-Meier survival plot was drawn to feature the relationship between TP53 mutations and survival of patients. As a complement to the previous analysis, a box plot is created to contrast the difference between the distribution of survival time of young patients with and without TP53 mutations. Lastly, I created a DESeqDataSet (dds) object with the RNA data using the DESeq2 library from a Bioconductor package and created a volcano plot to illustrate the regulation of genes in the young and old age groups using EnhancedVolcano package. The pathological stage factor is also included in the design of the dds object to minimize the effects of pathological stage as a confounding variable. Additionally, genes that have a total of less than ten counts are removed to minimize noise.

Results

The maf survival analysis and the Kaplan-Meier survival plot illustrate opposite results. However, both p-values, 0.239 and 0.17, are greater than 0.05, meaning that there is no statistically significant evidence and I fail to reject the null hypothesis that there is no difference between the survival probability between young breast cancer patients with and without mutations in TP53. In the box plot, the median and IQR of the survival time of young patients with TP53 mutations is somewhat lower than those of patients with no TP53 mutations. Nevertheless, the distribution of survival time for the group with mutations in TP53 is more right skewed, indicating that although the majority of patients with mutated TP53 has a shorter survival time, those patients with a high percentile in survival time (above 75th) tend to have a

longer survival time than patients with no TP53 mutations. In the volcano plot, there is no data point above the significance threshold, meaning that no gene from the data set has statistically different expression rates for the younger group and the older group. The log2FC for TP53 is about 0.146, so the approximate fold change would be 1.11, meaning that the counts of TP53 in young patients is about equivalent to the TP53 counts in old patients. The adjacent p-value for TP53 is about 0.059, which is above the significance value of 0.05. In conclusion, mutation in TP53 does not seem to play a vital role in the survival rate of young breast cancer patients, and TP53 is not significantly regulated in young patients compared to old patients.

Discussion

The result of this analysis does not appear to be very consistent with previous research. For instance, mutations in TP53 have been discovered to be correlated with poor outcomes in acute myeloid leukemia (AML), which lead to lower survival rates (Nicholas J. S. et al, 2020). A possible explanation for this discrepancy is that the sample size of the young patient population is only 186, which may be too small to reach any significant value. Also, the median age of patient samples in the previously mentioned research is 70 years old, an age that is a lot larger than the age of the population in this analysis. For future studies, it would be better to have a larger sample size of the desired population (< 45 years old).

Figures

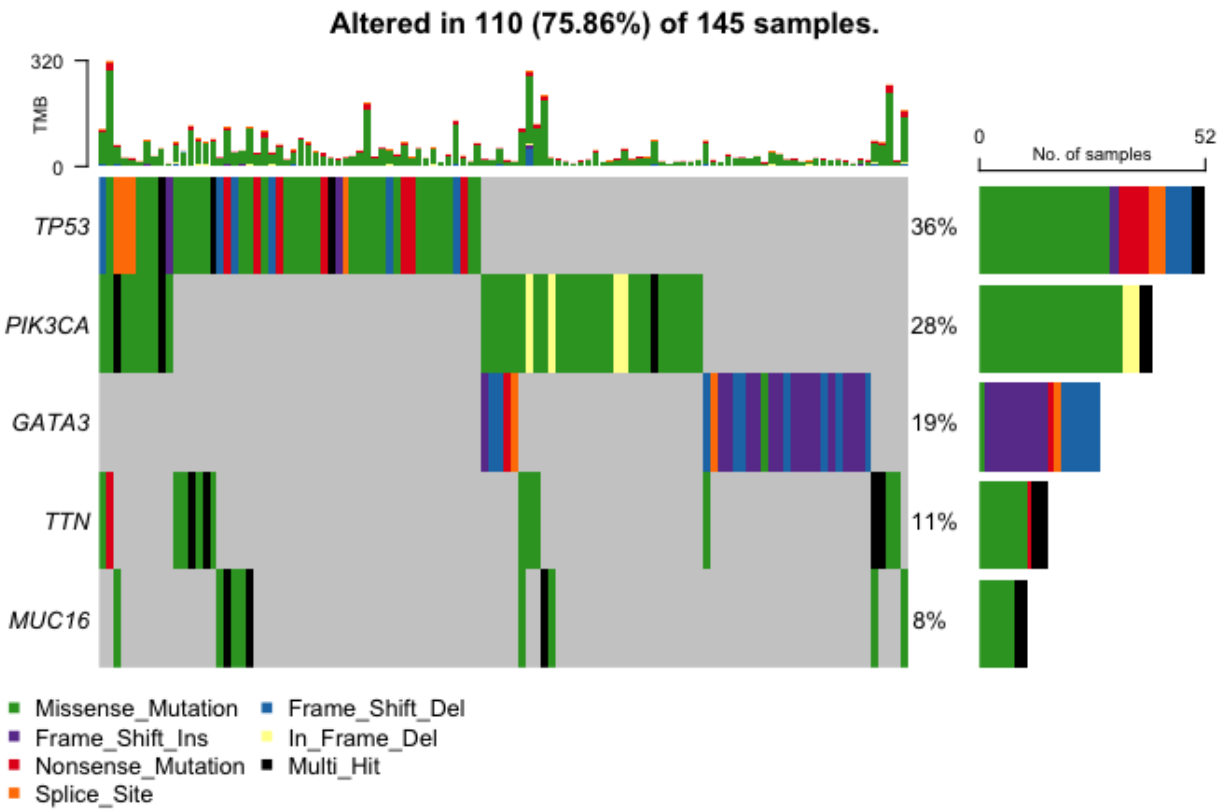

Fig. 1. Oncoplot showing the most common mutations in young breast cancer patients (<45 years old) are respectively TP53, PIK3CA, GATA3, TTN, and MUC16. The most common mutation type for TP53 is missense mutation.

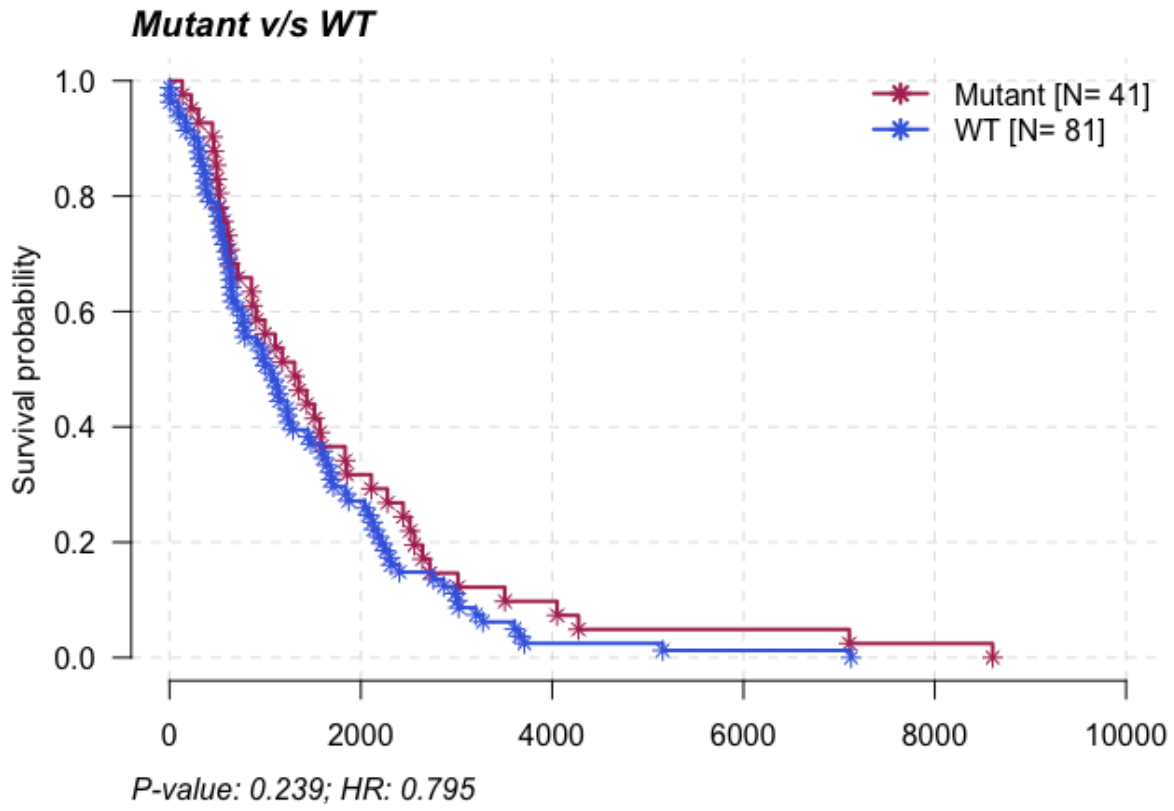

Fig. 2. Maf survival analysis shows the survival probability of cancer patients with mutated TP53 is slightly higher than those with the wild type over time. The p-value of 0.239 indicates that the difference is not statistically significant.

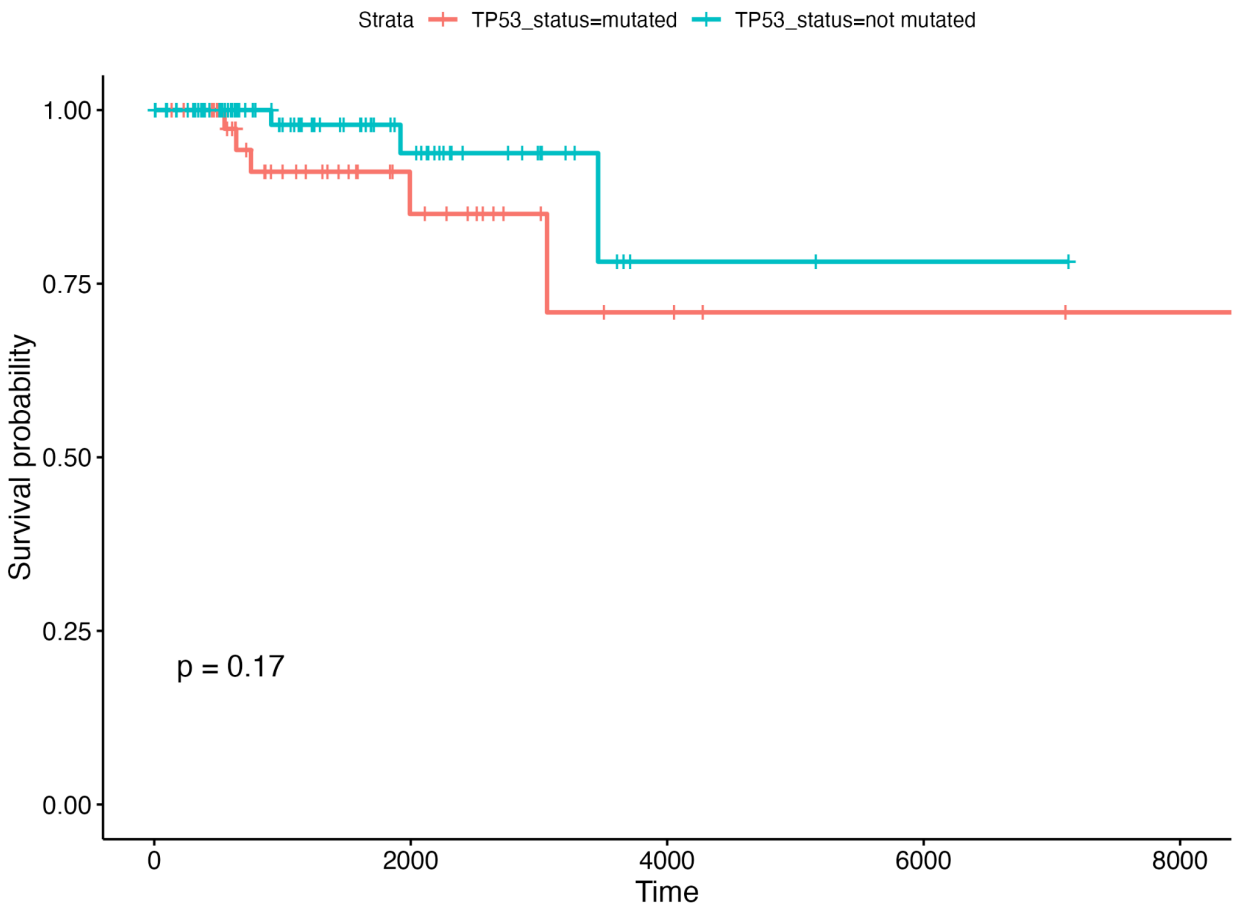

Fig. 3. Kaplan-Meier survival plot showing that young patients (<45 years old) with no TP53 mutation have a slightly higher survival probability than those with TP53 mutations. The p-value of 0.17 indicates that the difference is not statistically significant.

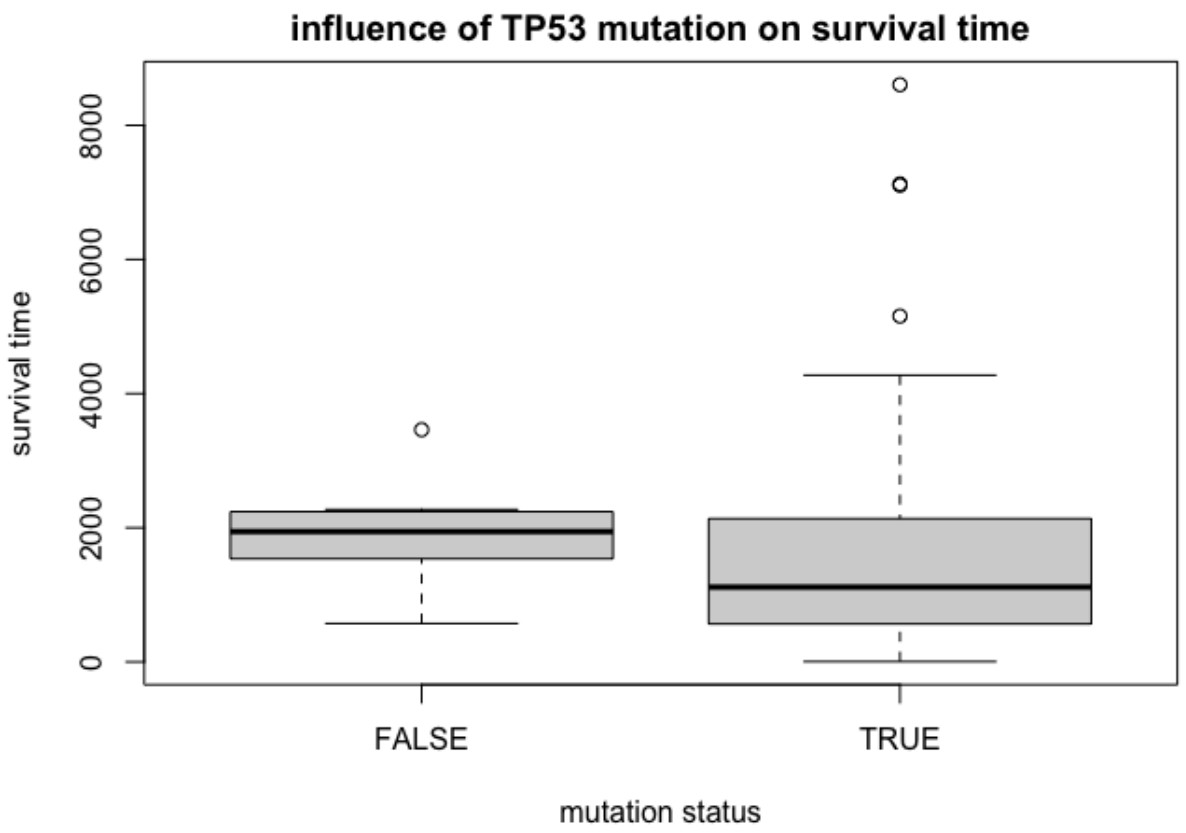

Fig. 4. Boxplot showing the center of the survival time for young patients (<45 years old) with no TP53 mutations is slightly longer than the central survival time of those with TP53 mutations. The survival time distribution for patients with TP53 mutations is more skewed toward right and has more high outliers.

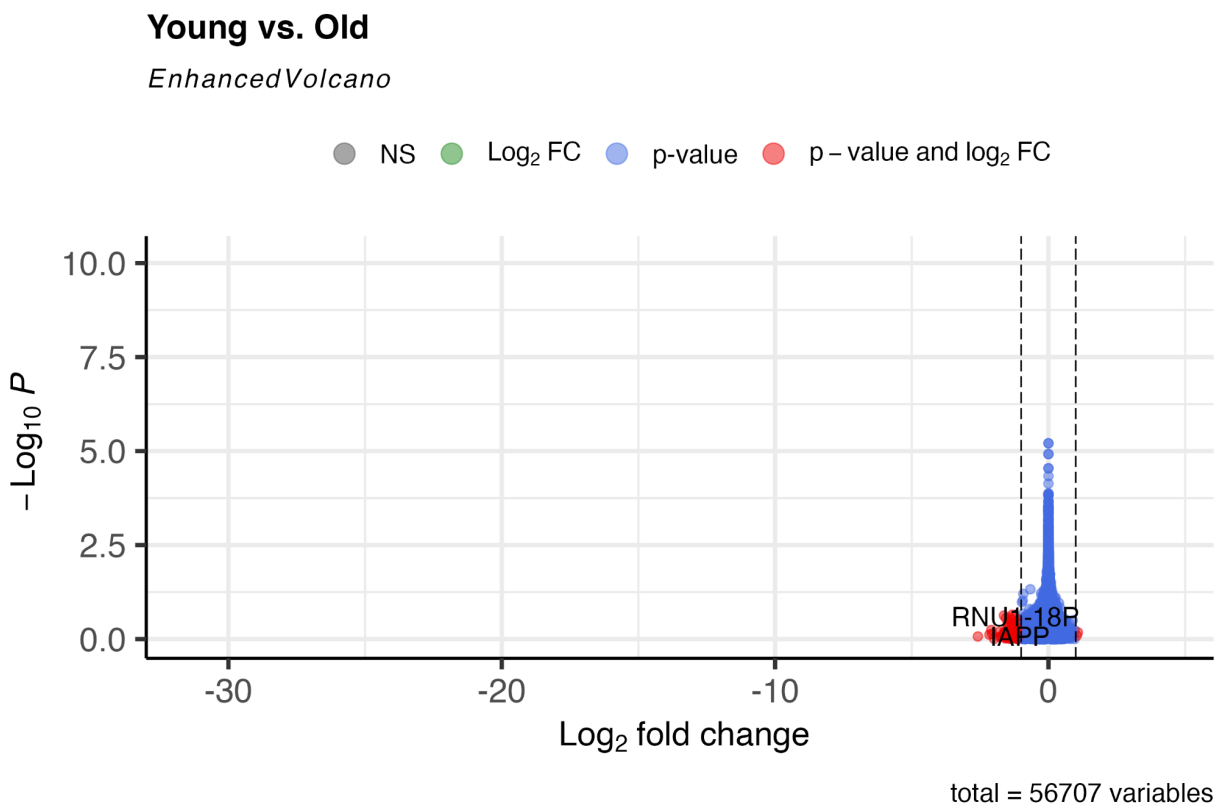

Fig. 5. Volcano plot showing the regulation of genes in young patients (<45 years old) in comparison to old patients (>= 45 years old). The potential confounding variable pathological stage is controlled. The p-value cut-off of the volcano plot is customized to 1.3 (-log10(0.05) = 1.3) because instead of adjacent p-value, the negative log 10 of adjacent p-value is used.

References

Akram, M., Iqbal, M., Daniyal, M. *et al.* Awareness and current knowledge of breast cancer. *Biol Res* 50, 33 (2017). https://doi.org/10.1186/s40659-017-0140-9

Short, N., Montalban-Bravo, G., Hwang, H., Ning, J. *et al*. Prognostic and therapeutic impacts of mutant *TP53* variant allelic frequency in newly diagnosed acute myeloid leukemia. *Blood Adv* November 24, 2020; 4 (22): 5681–5689. doi: https://doi.org/10.1182/bloodadvances.2020003120

Nie, G. (2017, May 30). *Breast cancer in "young" women: Age cutoff and clinicopathologic features.* An American Society of Clinical Oncology Journal. Retrieved March 9, 2023, from https://ascopubs.org/doi/abs/10.1200/JCO.2017.35.15_suppl.e12019

U.S. Cancer Statistics, Changes Over Time: Female Breast New Cancers, All Ages, All Races and Ethnicities, Female (n.d.). Retrieved from https://gis.cdc.gov/Cancer/USCS/#/Trends/.