

Surgical Procedures and the Characteristics of Breast Cancer

Introduction

To analyze breast cancer through a multi-omic lens, the TCGA dataset was used in this research. TCGA, or The Cancer Genome Atlas, provides public access to rich genomic and transcriptomic datasets surrounding 33 different cancer types from information of over 11,000 patients. This research paper will specifically focus on the genomics and transcriptomics surrounding breast cancer. The genomics will focus on the effects of DNA mutations while the transcriptomics will focus on the effects of RNA transcription. Breast cancer is the second most common type of cancer in women in the U.S. after skin cancer. The risk of getting breast cancer can be influenced by a variety of factors, including gender, age, and genetic mutations. For example, most breast cancers are found in older women (50 years or older) who have genetic mutations in the BRCA1 or BRCA2 gene. According to the CDC, in 2019, 264,121 new cases of female breast cancer were reported among women in the U.S., and 42,280 women died because of the cancer (CDC, 2022). Globally, in 2020, there were 685,000 deaths total. Fortunately, despite these severe consequences, countries that have succeeded in reducing breast cancer mortality have been able to achieve an annual breast cancer mortality reduction of 2-4% per year. If this annual mortality reduction can occur worldwide, then over 2.5 million breast cancer deaths can be avoided between 2020 and 2040 (WHO, 2021). This research paper seeks to address the correlation between surgical procedures on specific characteristic features of breast cancer, such as patient survival probability, genetic mutations, as well as other clinical factors. R and RStudio was used to compute and analyze the datasets, which includes creating the graphs shown in the Results section.

Methods

After reading in the necessary data, including the clinical data, DNA mutation data, and RNA sequencing data in R, the first step was to clean the data so that all NA values and “Other” values are excluded. To analyze how the surgical procedure affects patient survival, a Kaplan-Meier survival plot was made (Figure 1). A survival time column was created in the clinical data data frame with the non-numeric NA and -Inf values removed. A death event column was also created and the survival and fit objects initiated. All of these variables combined were used to create the final survival plot.

To analyze the relationship between genetic mutations and surgical procedure, a co-oncoplot was created to observe which genes are the most mutated in different surgical procedures (Figure 2). Because the previous survival plot displayed a significant difference in survival probability between patients who underwent a lumpectomy versus patients who underwent a simplified mastectomy, these two surgical procedures were specifically chosen for comparison in the co-oncoplot. To compare the number of genetic mutations against clinical data, the mutation data data frame was merged with the mutation data clinical data frame (which contains the same data as the original clinical data frame). Again, the surgical procedure column was cleared of NA and “Other” values and rewritten as a factor. Two 3D maf files were subsetted based on the patient’s surgical procedure (lumpectomy or simple mastectomy). The co-oncoplot was created with these two subsets. To look more closely at the most mutated gene for both surgical procedures, TP53, a co-lollipop plot was created to compare the mutations of this gene in lumpectomy patients vs. simple mastectomy patients (Figure 3).

To analyze the correlation of the type of surgical procedure to another clinical variable like the age of the patient, a contingency table and corresponding mosaic plot was created (Figure 4). The age of the patient was divided into two categories, young versus old, with young patients characterized as being at most 58 years old. There were again two types of surgical procedures that were being compared, lumpectomy and simple mastectomy. A Fisher test was then run on the contingency table. For a deeper investigation into the gene regulation patterns of breast cancer, a volcano plot was created (Figure 5). DESeq2 was run with the race of the patient and the pathological stage of the cancer as covariables. The RNA count and RNA genes data frames were filtered so that all the genes with the total number of counts less than 10 across all patients are excluded. The resulting dataset was converted into a dataframe with the unnecessary rows and columns eliminated and with the columns renamed to label the dataframe in a clearer way. An enhanced volcano plot was created using this dataframe. In addition, two new data frames were created to represent upregulated genes and downregulated genes by sorting the original dataframe in descending and ascending order of log2 fold change values respectively. The genes in these data frames were compared to the genes shown in the co-oncoplot.

Results

Based on Figure 1, patients who underwent simple mastectomy have a significantly higher survival probability than those who underwent lumpectomy and modified radical mastectomy. From time 0-2000, all three survival plots are seen at around the same survival probability range. At around time 2000, the survival probability starts rapidly decreasing for modified radical mastectomy patients, while at around time 3000, the survival probability starts rapidly decreasing

for lumpectomy patients. The difference in survival probability between simple mastectomy patients and lumpectomy patients at this point is around 0.25.

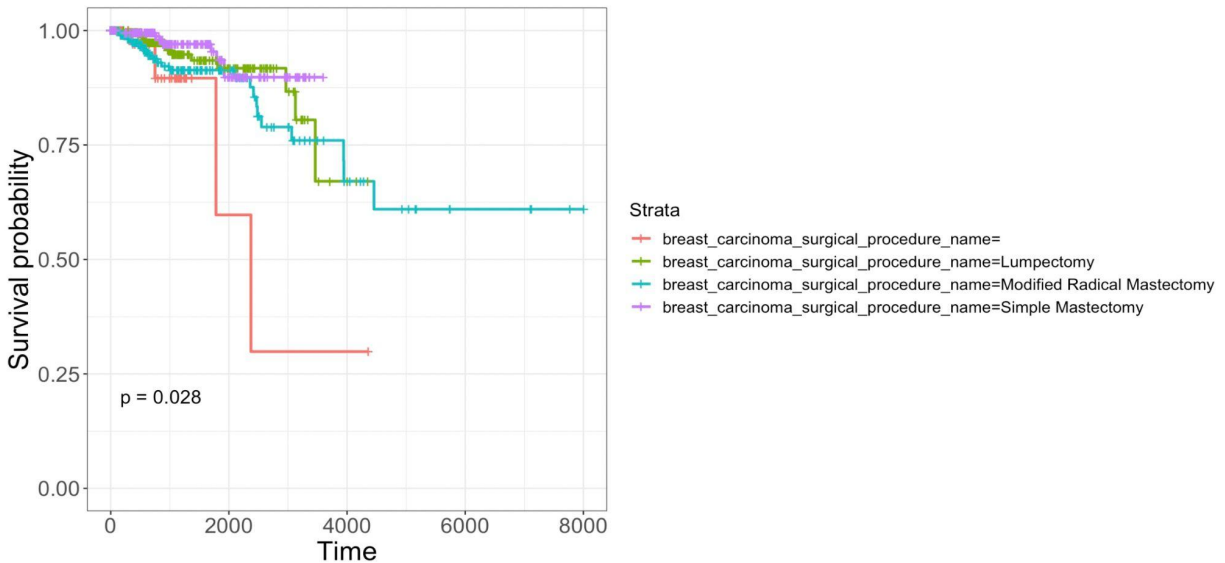


Figure 1. A Kaplan-Meier survival plot comparing the survival probability of patients over time for each type of surgical procedure.

Figure 2 shows that TP53 is the most highly mutated gene in lumpectomy patients. 44% of lumpectomy patients have this gene mutation as compared to only 23% of simple mastectomy patients. Another gene to take note of is CDH1, in which there are more mutations in patients who underwent the simple mastectomy surgery (21%) as compared to those who underwent the lumpectomy surgery (9%).

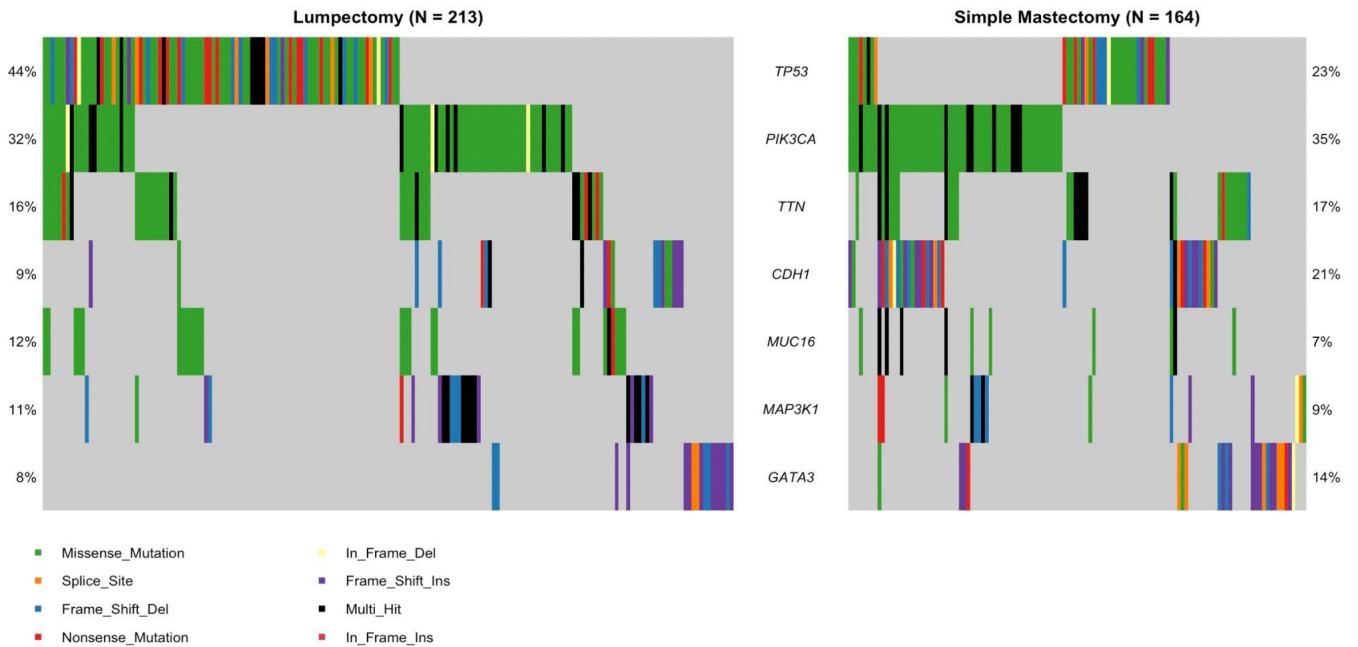


Figure 2. A co-oncoplot comparing the top 7 most mutated genes between patients who underwent lumpectomy (N = 213) and patients who underwent simple mastectomy (N = 164) with the top most mutated gene listed at the top. The different colors represent the type of mutation and each column represents a patient.

Within the TP53 gene, there are significantly more missense mutations in lumpectomy patients as compared to simple mastectomy patients (Figure 3). There are also some more nonsense mutations in lumpectomy patients. The location of the mutations are similar between the two groups, with slightly more mutations in the P53_tetramer domain in lumpectomy patients.

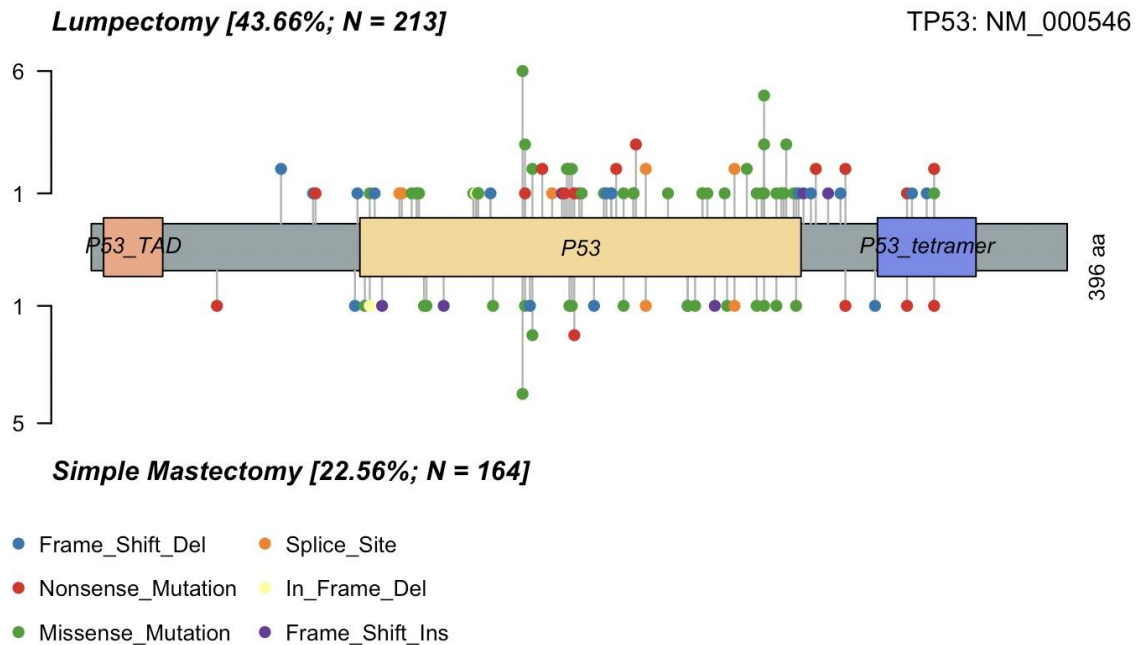


Figure 3. A co-lollipop plot comparing TP53 gene mutations between patients who underwent lumpectomy and patients who underwent simple mastectomy. 43.66% of lumpectomy patients had TP53 gene mutations as compared to only 22.56% of simple mastectomy patients. The different colored “lollipops” represent the type of mutation and the different colored boxes represent the gene domain. The height of each “lollipop” represents the number of patients with the gene mutation.

The results of a Fisher test show that there is a 1.49 times chance of being an older patient if a lumpectomy surgery was performed as compared to a simple mastectomy. These results are significant, having a p-value of 0.03869, which is less than a significance level of 0.05. The mosaic plot in Figure 4 shows further support for this, as the area of the boxes indicate that older patients are more likely to get lumpectomy surgeries as compared to a simple mastectomy while younger patients are more likely to get simple mastectomy surgeries as compared to a lumpectomy.

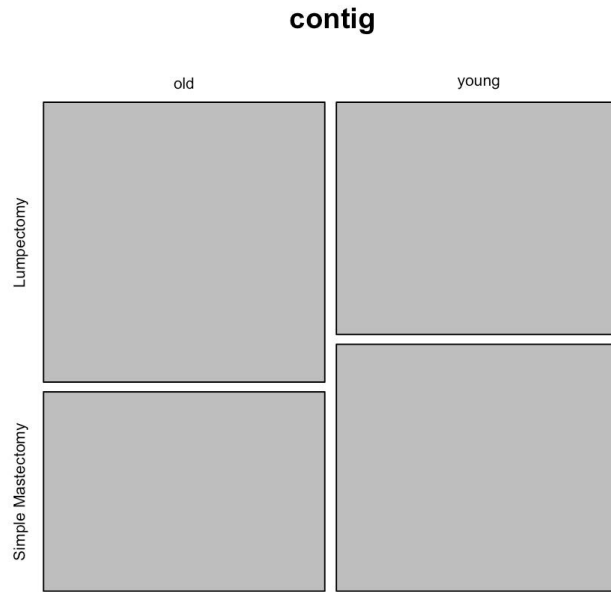


Figure 4. A mosaic plot comparing two categorical groups of two variables: age (old vs. young) and surgical procedure (lumpectomy vs. simple mastectomy) with the area representing each proportion. (Fisher test: p-value = .03869, odds ratio = 1.491453)

Some genes that are significantly upregulated in younger patients are TNFS11 and RELN, while genes that are significantly downregulated in younger patients are ACTG1P22 and LINC02551. None of the highly mutated genes, such as TP53, found from Figure 2 are part of the significantly upregulated and downregulated genes shown in Figure 5.

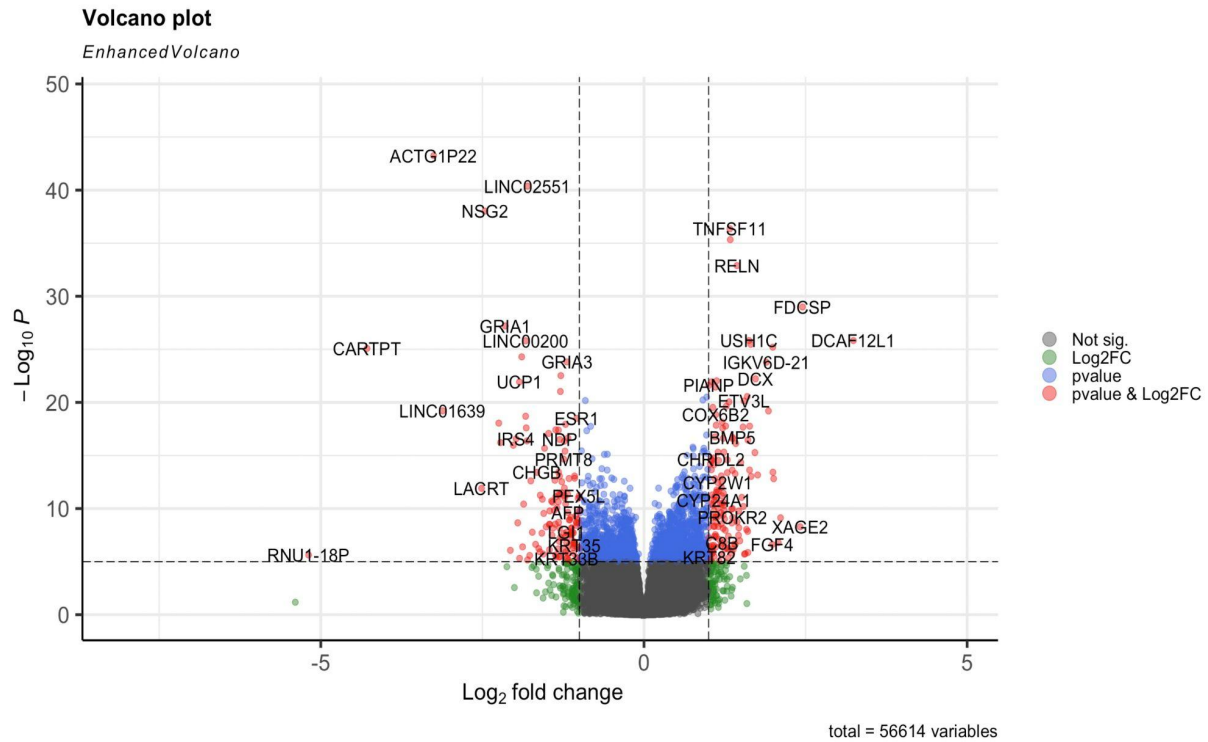


Figure 5. A volcano plot comparing the age of patients across all genes, showing whether each gene is upregulated or downregulated for young (≤ 58 years) and old patients (> 58 years).

Discussion

For this research, two main surgical procedures for breast cancer were compared, lumpectomy and simple mastectomy. A lumpectomy only removes the cancerous and abnormal tissue from the breast, while a simple mastectomy removes the whole entire breast. Although simple mastectomy seems to be more dangerous, it was shown in Figure 1 that there is a significantly higher survival probability in patients who undergo this surgery as compared to other surgeries. Previous studies have also pointed out that lumpectomy is associated with a higher 20-year mortality than mastectomy (Laurberg et al., 2016). Additionally, mastectomy is associated with less chronic pain frequency and less post-surgical side effects than lumpectomy (Czajka &

Pfeifer, 2022). The difference in survival probability shown in the KM plot could potentially be due to these factors. Furthermore, the highest occurring gene mutation in lumpectomy patients was the TP53 gene mutation (Figure 2), which occurred at almost twice the occurrence of simple mastectomy patients and was composed mostly of missense and nonsense mutations (Figure 3). This actually contradicts existing literature, as mastectomy is recommended for patients who exhibit mutations in the TP53 gene (Schon & Tischkowitz, 2017). This is to avoid adjuvant breast radiotherapy as TP53 gene mutation carriers (with TP53 being a tumor suppressor gene, are more susceptible to the reformation of tumors). The difference in sample size between simple mastectomy patients and lumpectomy patients could be an explanation for this contradiction.

The mosaic plot (Figure 4) showed that there is a slightly higher chance for older patients to receive a lumpectomy and younger patients to receive a simple mastectomy. This corresponds to the findings from the study conducted by Laurberg et al. that found that younger women who had lumpectomy were 60-70% more likely to die from breast cancer 20 years after diagnosis than older women who had lumpectomy (2016). Younger women would therefore have a higher chance of survival by undergoing a different surgical procedure, such as simple mastectomy, supporting the discrepancy seen in the mosaic plot. However, the significant genes shown in Figure 5 don't match any of the genes shown in Figure 2. Therefore, although age could play a role in determining which surgical procedure can lead to the best health outcome, making this decision based on gene regulation isn't an effective method. The surgical procedure a patient undergoes can be correlated with both genetic mutations and age, and age can be correlated with certain patterns of gene regulation, but gene regulation and the type of surgical procedure don't seem to be correlated with each other.

There were some limitations to the research conducted, since the sample sizes of lumpectomy patients compared to simple mastectomy patients were notably different ($N = 213$ vs. $N = 164$). This can lead to discrepancies that aren't necessarily caused by the factors mentioned above, but simply due to not having enough patient data or having access to data of a greater number of patients. Nevertheless, this research paper uncovers significant findings regarding the relationship of a patient's surgical procedure with other breast cancer characteristics that can help determine which surgery would be best for a patient. Future research can consist of further analysis of other clinical factors besides age that are correlated with the type of surgical procedure a patient undergoes, such as the presence of certain receptors related to breast cancer, race, or the pathological stage of cancer. The CDH1 gene could also be more closely examined to help explain the difference in its mutation rates between lumpectomy and simple mastectomy patients.

References

- Admoun, C., & Mayrovitz, H. (2021). Choosing Mastectomy vs. Lumpectomy-With-Radiation: Experiences of Breast Cancer Survivors. *Cureus*, 13(10), e18433.
<https://doi.org/10.7759/cureus.18433>
- Centers for Disease Control and Prevention. (2022, September 26). Breast cancer. Centers for Disease Control and Prevention. Retrieved March 11, 2023, from
<https://www.cdc.gov/cancer/breast/index.htm>
- Czajka, M. L., & Pfeifer, C. (2022, September 12). Breast cancer surgery - statpearls - NCBI bookshelf. National Library of Medicine. Retrieved March 12, 2023, from
<https://www.ncbi.nlm.nih.gov/books/NBK553076/>
- Laurberg, T., Lyngholm, C. D., Christiansen, P., Alsner, J., & Overgaard, J. (2016). Long-term age-dependent failure pattern after breast-conserving therapy or mastectomy among Danish lymph-node-negative breast cancer patients. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*, 120(1), 98–106.
<https://doi.org/10.1016/j.radonc.2016.05.006>
- Schon, K., & Tischkowitz, M. (2018). Clinical implications of germline mutations in breast cancer: TP53. *Breast cancer research and treatment*, 167(2), 417–423.
<https://doi.org/10.1007/s10549-017-4531-y>
- World Health Organization. (2021, March 26). *Breast cancer*. World Health Organization. Retrieved March 11, 2023, from
<https://www.who.int/news-room/fact-sheets/detail/breast-cancer>