Introduction:

The Cancer Genome Atlas, TCGA, is a comprehensive cancer genomics program that is used to accelerate public understanding of cancer on a molecular level. TCGA is a prominent cancer database because it makes comparisons between normal tissues and 30 different cancer types. TCGA is particularly important because it is able to look at genetic alterations via DNA sequencing, RNA sequencing, and epigenetic data. In addition to the genomic data, TCGA also is noted to include information patient data for each of the patients are included in the dataset. This comes in the form of patient demographics, survival data, and clinical history of each patient. Though this data is not exhaustive across all the patients included in the sample, the large size of the dataset provides a large enough sample size to look within a specific cancer type. This study will specifically investigate breast cancer. Breast cancer can occur in three different areas of the breast which includes the lobules, ducts, and connective tissue. Lobules contain the glands that produce milk while ducts help to carry milk to the nipple (Boughey et al., 2013). Connective tissue helps to maintain structure in the region. The most common breast cancers are invasive ductal carcinomas or invasive lobular carcinomas which are cancers in the ducts or lobules respectively. The common diagnosis pathway starts with breast ultrasound imaging and diagnostic mammograms which are routinely recommended every two years for patients over the age of 50 (Kunkler et al., 2023). If mammograms begin to detect abnormal growths, providers often recommend retrieval of a solid biopsy from the regions of the breast where the tumor is detected. If determined to be breast cancer, a common treatment option would include surgical intervention. The goal of this intervention is to remove as much of the cancerous tumor as possible. This can be accomplished through both breast-conserving or complete mastectomy depending on the anatomical positioning of the tumor. In addition to the anatomical positioning

of the tumor, it is essential to look at adjacent lymph nodes to see if spread has taken place across the lymphatic system in order to determine the state of metastasis.

Keeping these in mind, surgical intervention can vary based on growth and the physician's assessment of the tumor. Subjectivity in this area provides opportunities to make adjustments to surgical interventions. The goal of this study is to use TCGA's breast cancer dataset in order to observe how survival is impacted based on the variety of surgical interventions performed on breast cancer patients. Specific emphasis will be placed on the differential expression profile on patients with mutations on the BRCA1 and BRCA2 genes as these are the most commonly found mutations in patients with breast cancer. BRCA1 is implicated in DNA repair while BRCA2 is implicated in suppression of cell growth (Toesca et al., 2021). These two genes work in order to suppress tumor growth when functioning normally. This study will also look at whether BRCA1 and BRCA2 are co-occurring in their function and whether surgical intervention could function.

Methods:

This project started by installing the necessary packages on RStudio. This include installation of survival, survminer, ggplot2, and maftools which helped with the initial visualizations of TCGA data. With the outcome metric being survival, it was necessary to create a new column on the TCGA database that calculated individual patient survival. In this process, patients with non-numerical values were removed. A binary classification was used as the outcome event that would differentiate patients based on result from surgical intervention. This was classified as a "true" or "false" with true indicating that the patient has died and false indicating that the patient was known to be living at the latest data collection. Patients that did not have surgical procedure were changed to "No Procedure" which allowed for further differentiation of the effectiveness of surgical intervention. Patients with infinite values were also removed from the study. This

allowed for generation of a KM plot through a survival analysis. The next step was seeing how this compared with BRCA1 and BRCA2 mutations in patients with breast cancer. An oncoplot was used to observe patterns of genetic alterations in BRCA1 and BRCA2. The coOncoplot was used to see whether these genetic alterations were mutually exclusive with each other. This was followed by two lollipop plots for BRCA1 and BRCA2 respectively and compared patients that underwent surgery and patients who did not go through surgery. A volcano plot was used to determine whether BRCA1 and BRCA2 were significantly upregulated or downregulated in expression.

Results:

This study included patients who varied based on surgical procedure and patients who did not undergo surgical procedure. The distribution of patients and their procedure can be seen in Figure 1.

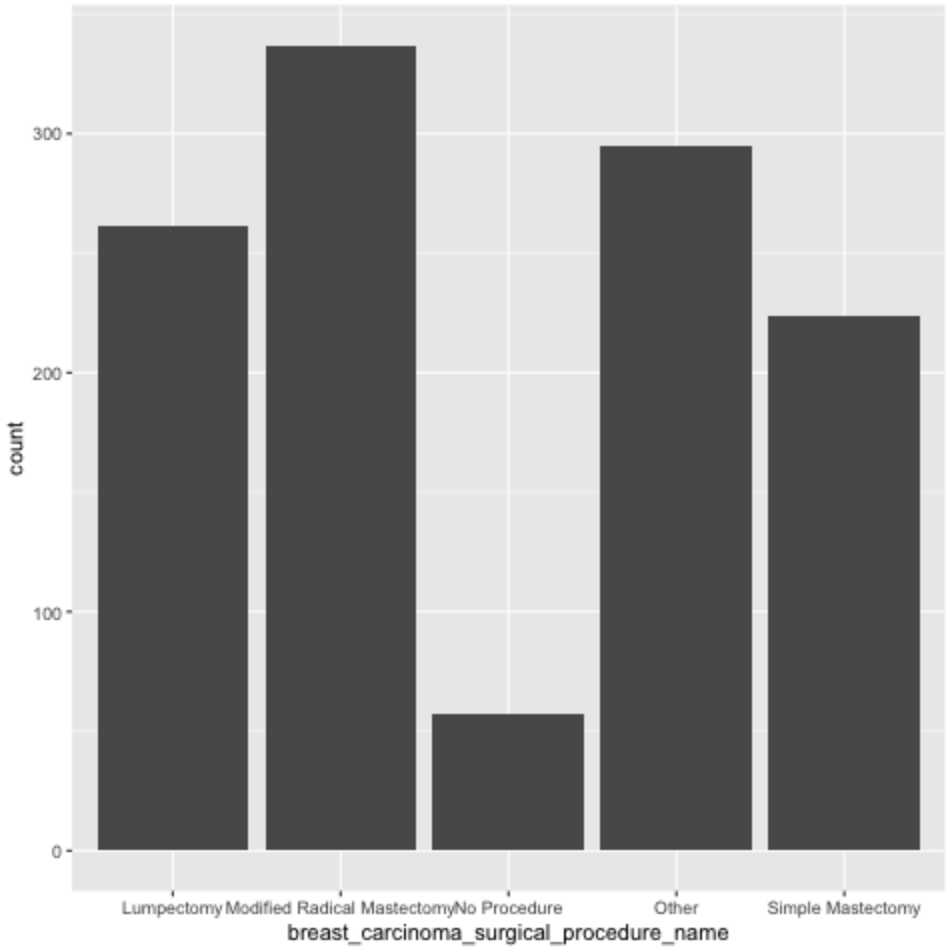

Figure 1: Surgical Procedure Distribution

This is followed by a survival analysis plot that was developed based on the different surgery types.

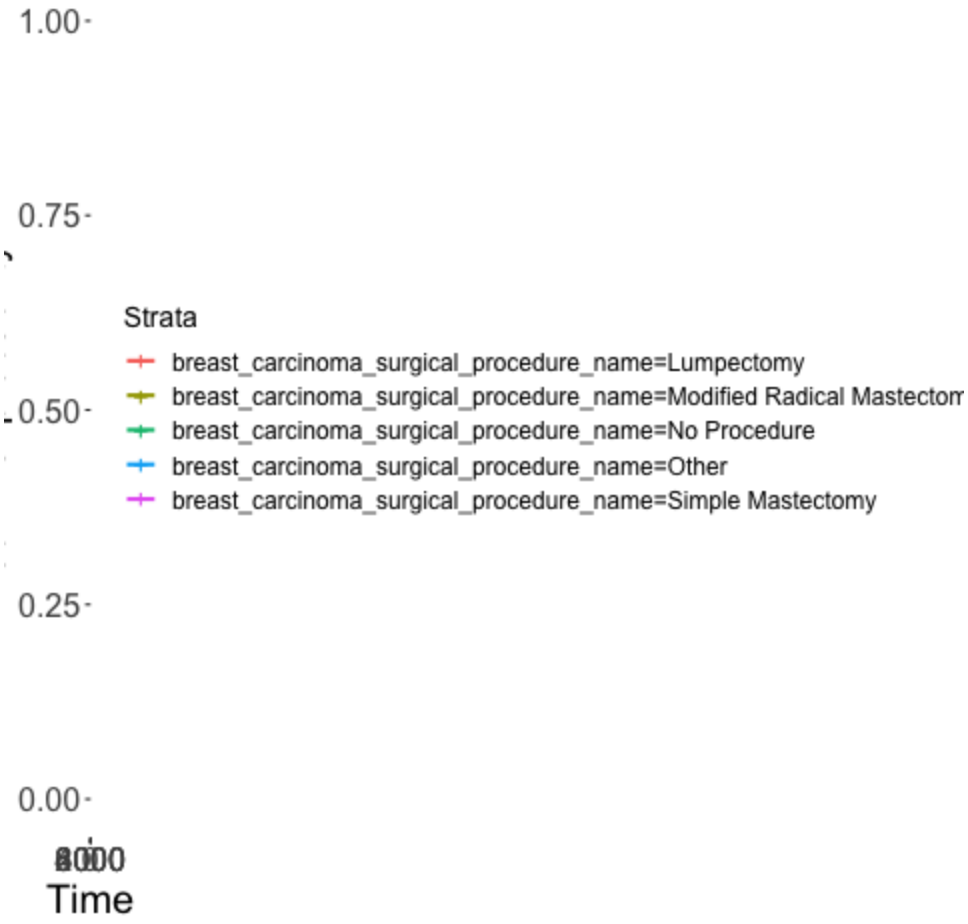

Figure 2: Survival Analysis Kaplan-Meier Plot based on Surgical Intervention

This was followed by a differential analysis of BRCA1 and BRCA2 mutations in patients.

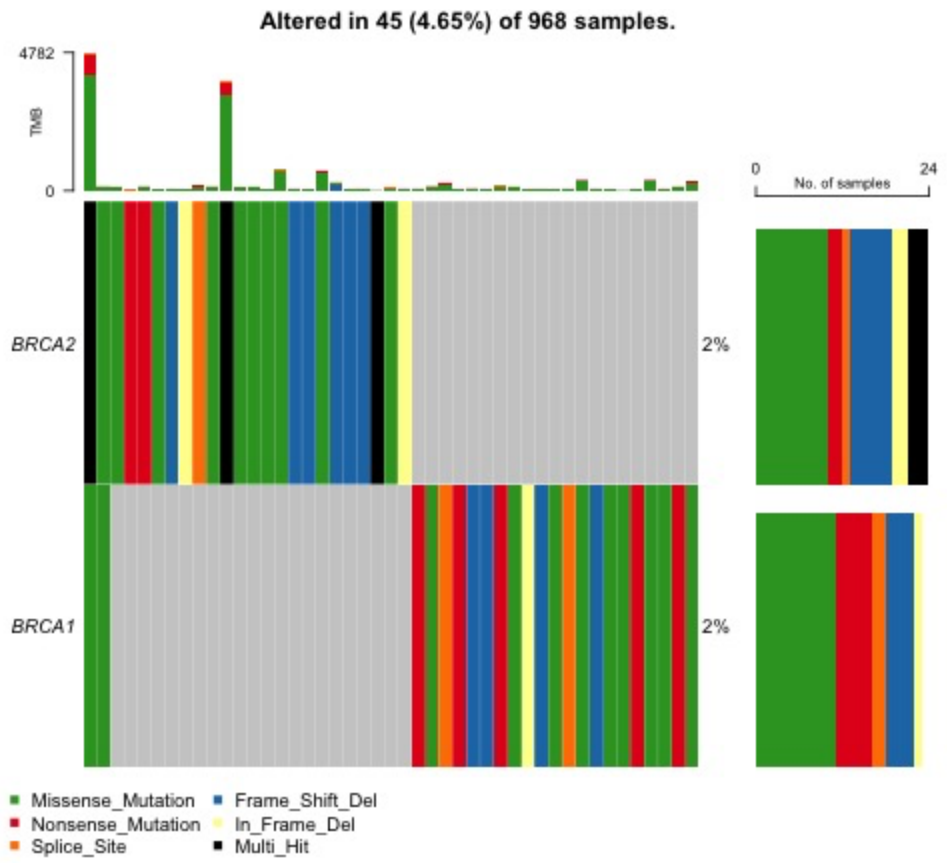

Figure 3: Oncoplot for BRCA1 and BRCA2

The coOncoplot helps to see how if BRCA1 and BRCA2 mutations are co-occurring or display a relationship.

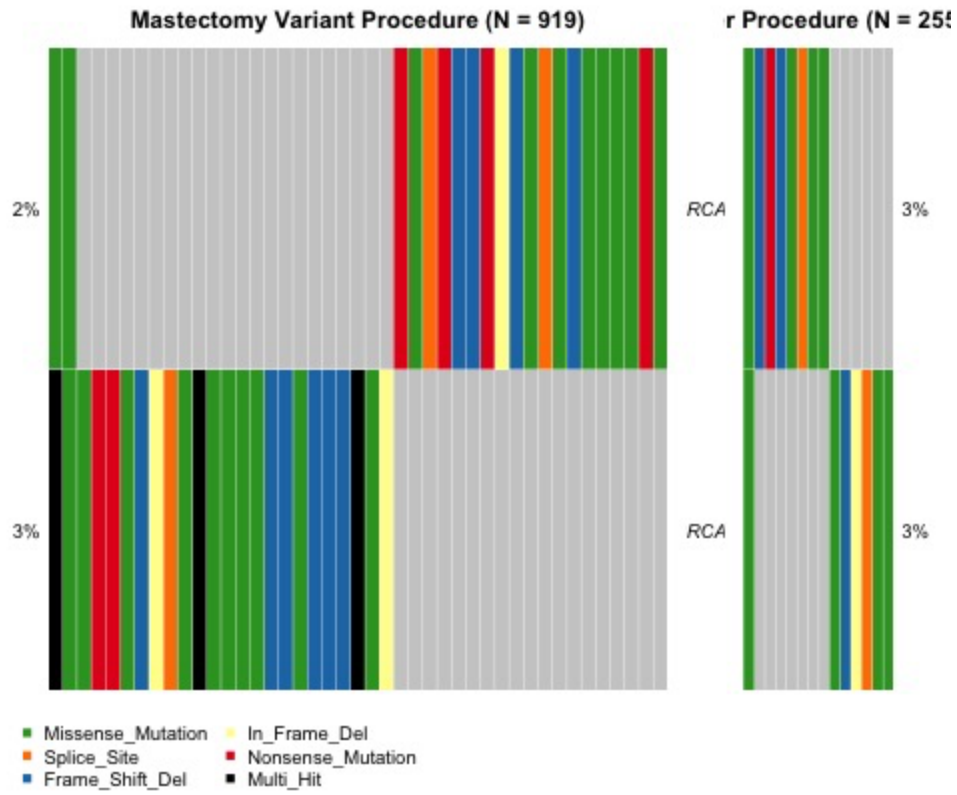

Figure 4: coOncoplot for BRCA1 and BRCA2

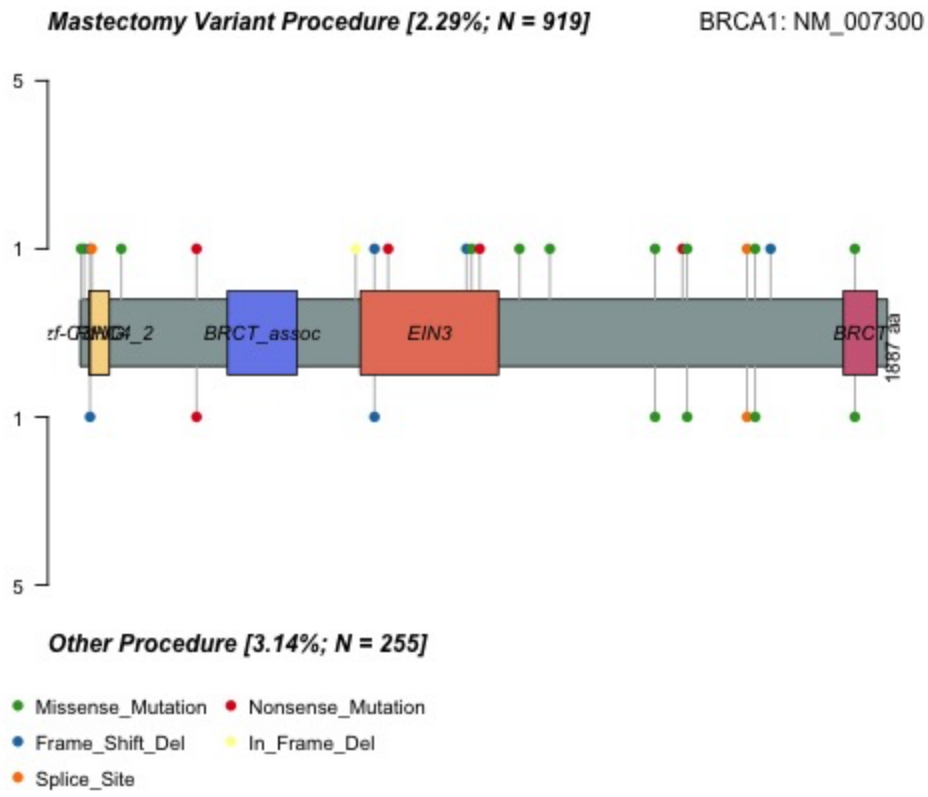

Figure 5: Lollipop plot for BRCA1 based on surgery or no surgery

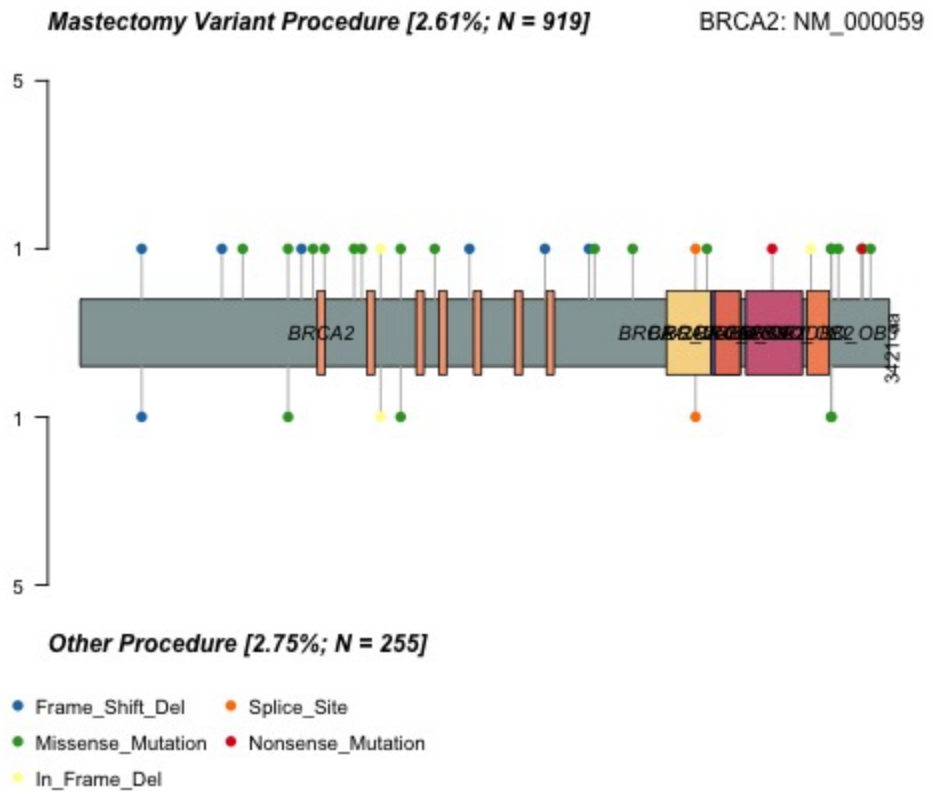

Figure 6: Lollipop plot for BRCA2 based on surgery or no surgery

Discussion:

From Figure 1, we can see a clear skew towards patients who have undergone surgery in comparison to patients who did not. The majority of patients appear to have undergone a radical mastectomy which is likely to be the most aggressive procedure for the removal of a tumor. With regards to gene expression, it appears that BRCA1 and BRCA2 are not co-occurring in breast cancer and appear to not have a defined relationship. This is shown in Figure 3 and 4 which appears to show mutations appearing in two different areas of the gene. Thus, it is possible to speculate that the effects of BRCA1 and BRCA2 are independent of each other. Figures 3 and 4 also appear to show that a majority of mutations in BRCA1 and BRCA2 are missense mutations

that render the expression of the gene deleterious in nature. Figure 5 shows mutations varying throughout the gene. However, Figure 6 shows a larger number of missense mutations clustering throughout the gene, implying that changes to the amino acid sequence are likely rendering the protein product deleterious and preventing BRCA2 to be functioning normally.

References:

Boughey, J. C., Suman, V. J., Mittendorf, E. A., Ahrendt, G. M., Wilke, L. G., Taback, B., Leitch, A. M., Kuerer, H. M., Bowling, M., Flippo-Morton, T. S., Byrd, D. R., Ollila, D. W., Julian, T. B., McLaughlin, S. A., McCall, L., Symmans, W. F., Le-Petross, H. T., Haffty, B. G., Buchholz, T. A., Nelson, H., … Alliance for Clinical Trials in Oncology (2013). Sentinel lymph node surgery after neoadjuvant chemotherapy in patients with node-positive breast cancer: the ACOSOG Z1071 (Alliance) clinical trial. *JAMA*, *310*(14), 1455–1461. https://doi.org/10.1001/jama.2013.278932

Kunkler, I. H., Williams, L. J., Jack, W. J. L., Cameron, D. A., & Dixon, J. M. (2023). Breast-Conserving Surgery with or without Irradiation in Early Breast Cancer. *The New England journal of medicine*, *388*(7), 585–594. https://doi.org/10.1056/NEJMoa2207586

Toesca, A., Sangalli, C., Maisonneuve, P., Massari, G., Girardi, A., Baker, J. L., Lissidini, G., Invento, A., Farante, G., Corso, G., Rietjens, M., Peradze, N., Gottardi, A., Magnoni, F., Bottiglieri, L., Lazzeroni, M., Montagna, E., Labo, P., Orecchia, R., Galimberti, V., … Veronesi, P. (2022). A Randomized Trial of Robotic Mastectomy Versus Open Surgery in Women With Breast Cancer or BrCA Mutation. *Annals of surgery*, *276*(1), 11–19. https://doi.org/10.1097/SLA.0000000000004969