I. Introduction

TCGA–or–The Cancer Genome Atlas is an online database with support from the NIH that holds genomic, transcriptomic, epigenomic, and proteomic data for 33 different types of cancers that is publicly available for study. The different omic levels of data available in TCGA allow for powerful multi-omic data research. Multi-omic data involves the combination of different levels of the central dogma of molecular biology in order to get a more complete picture of biochemical systems. We used multi-omic TCGA breast cancer data to examine breast cancer more in-depth. Breast cancer itself is an incredibly deadly disease with it being a leading cause of cancer-related deaths in women worldwide and has been known to be affected by age, gender, family history, and lifestyle factors such as alcohol usage and exercise (Łukasiewicz). In our research looking at TCGA data, we used transcriptomic and genomic data involving Breast Invasive Carcinoma to look at radiation therapy efficacy. We wanted to know whether age affected the efficacy of radiation treatment, whether radiation patients had a better prognosis than those that did not receive radiation therapy, the commonly mutated genes that radiation patients have, and over/underexpressed genes that radiation patients have. We explored this through R using clinical data plotting, analysis of nonsilent SNPs in MAF files, and differential expression analysis. The research is designed to help physicians better understand when to administer radiation therapy and see whether it is effective at all. We found that radiation therapy was effective and that certain biomarkers such as CDH1 and RUNX1 can be indicative of radiation-resistant cancer while MUC16 could be a marker of breast cancer treatable with radiation therapy.

II. Methods

Data analysis was conducted in R Studio using the language R. In addition to the base R functions, Bioconductor, MAFtools, ggplot2, and TCGAbiolinks packages were imported. From the publicly available TCGA-BRCA (The Cancer Genome Atlas Breast Invasive Carcinoma) dataset clinical data, clinical radiation data, nonsilent SNP mutation (MAF) data, and transcriptomic data looking at RNA counts in the cancerous tissue were utilized to look at factors that might affect BRCA cancer. The clinical radiation and clinical data dataframes were combined together into one dataframe by patient in order to form a more complete look at patients' clinical history. From the combined dataframe, a simple bar graph was drawn using ggplot2 functions to compare the ages of diagnosis and radiation treatment outcomes. A Kaplan-Meier analysis was conducted using ggplot2 on the combined radiation-clinical dataframe and a Kaplan-Meier plot was produced. An oncoplot for the different levels of radiation therapy was generated using MAFtools and the nonsilent SNP MAF. A differential expression analysis of RNA counts was conducted using Bioconductor and the transcriptomic dataset and the results were translated into a volcano plot.

III. Results

First, we wanted to examine the correlation between the age of cancer diagnosis with the effectiveness of therapies. It is known that older patients react differently to many therapies than expected as well as data about older patients is scarce, so we wanted to look at the effectiveness of radiation therapy in curing older patients' breast cancer (Fialová et. al).

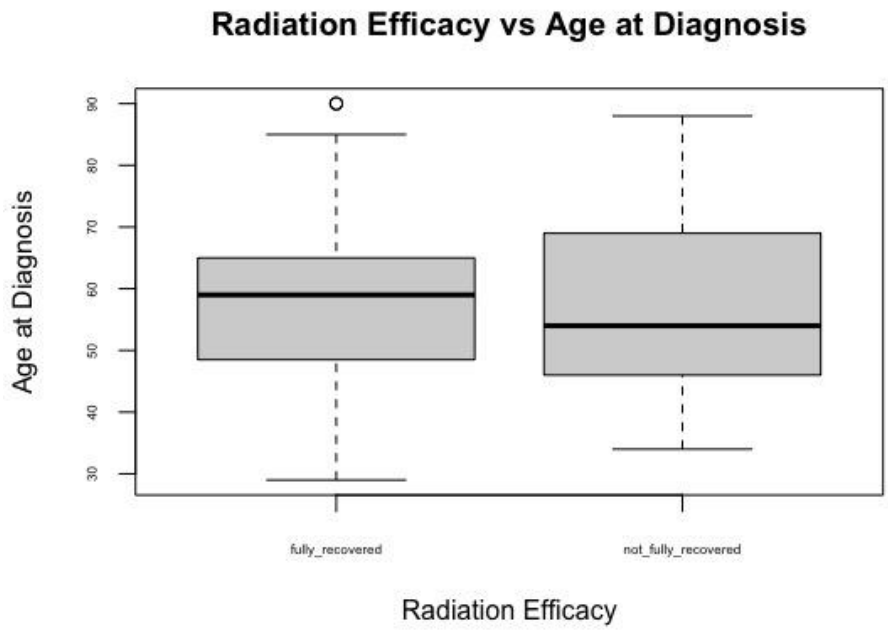

Figure 1. Boxplot comparing patients that fully recovered after radiation treatment (fullt_recovered) and those that did not (not_fully_recoevred) using their age. The fully recovered group has a higher median age than the not fully recovered group with one outlier present.

We concluded that the average age of diagnosis and, in turn, the age of the patient doesn't significantly change the efficacy of radiation therapy (Figure 1). However, the median age of patients that fully recovered was higher than the median age of patients that didn't fully recover which was an unexpected result (Figure 1). We then decided to conduct a Kaplan-Meier analysis to examine the survival outcomes of patients that received radiation therapy and fully recovered, those that didn't fully recover, and those patients that didn't receive radiation therapy.

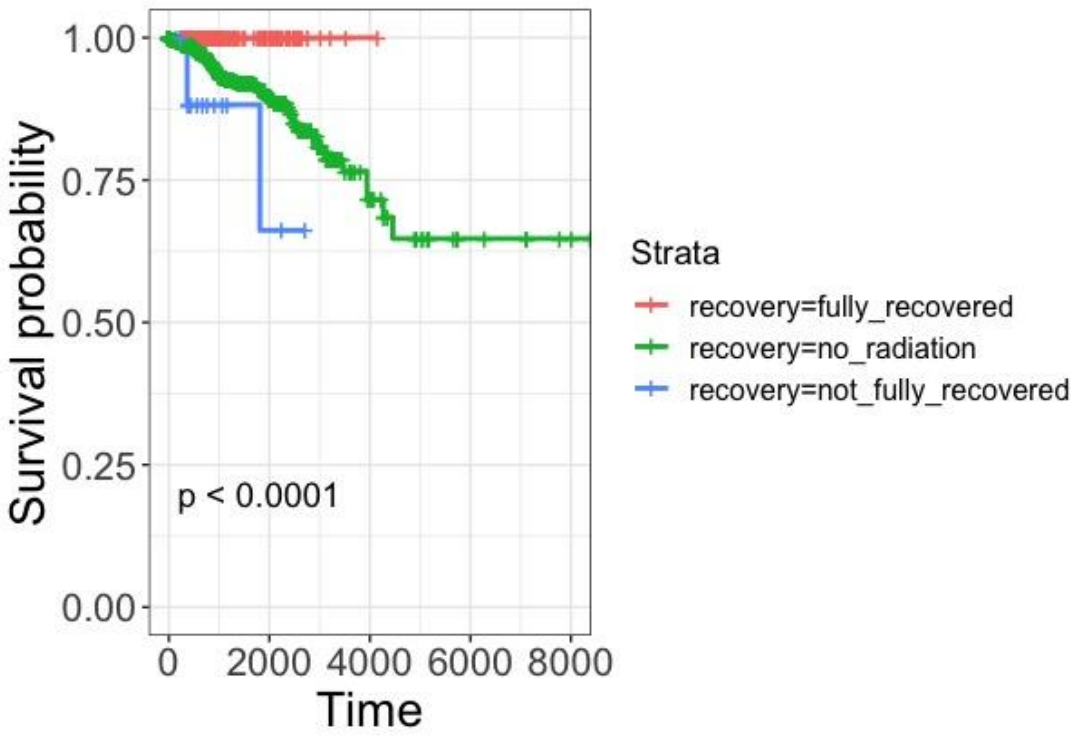

Figure 2. Kaplan-Meier plot of patients that had radiation therapy and fully recovered (fully_recovered), had radiation therapy and didn't fully recover (not_fully_recovered), and those that didn't have radiation therapy (not_fully_recovered). The best outcomes were for patients that fully recovered after radiation therapy and the worst outcomes were for patients that didn't fully recover after radiation therapy with the non-radiation therapy group having an outcome in between those two. The p-value of <0.0001 shows the significance of the results.

We found that patients that received radiation therapy and fully recovered from cancer had a better survival outcome than patients that didn't get treatment or had alternative treatments (Figure 2). However, patients that received radiation therapy and didn't fully recover had worse outcomes than either of the two groups (Figure 2). We then created oncoplots of the three groups as the not fully recovered group had a small sample size.

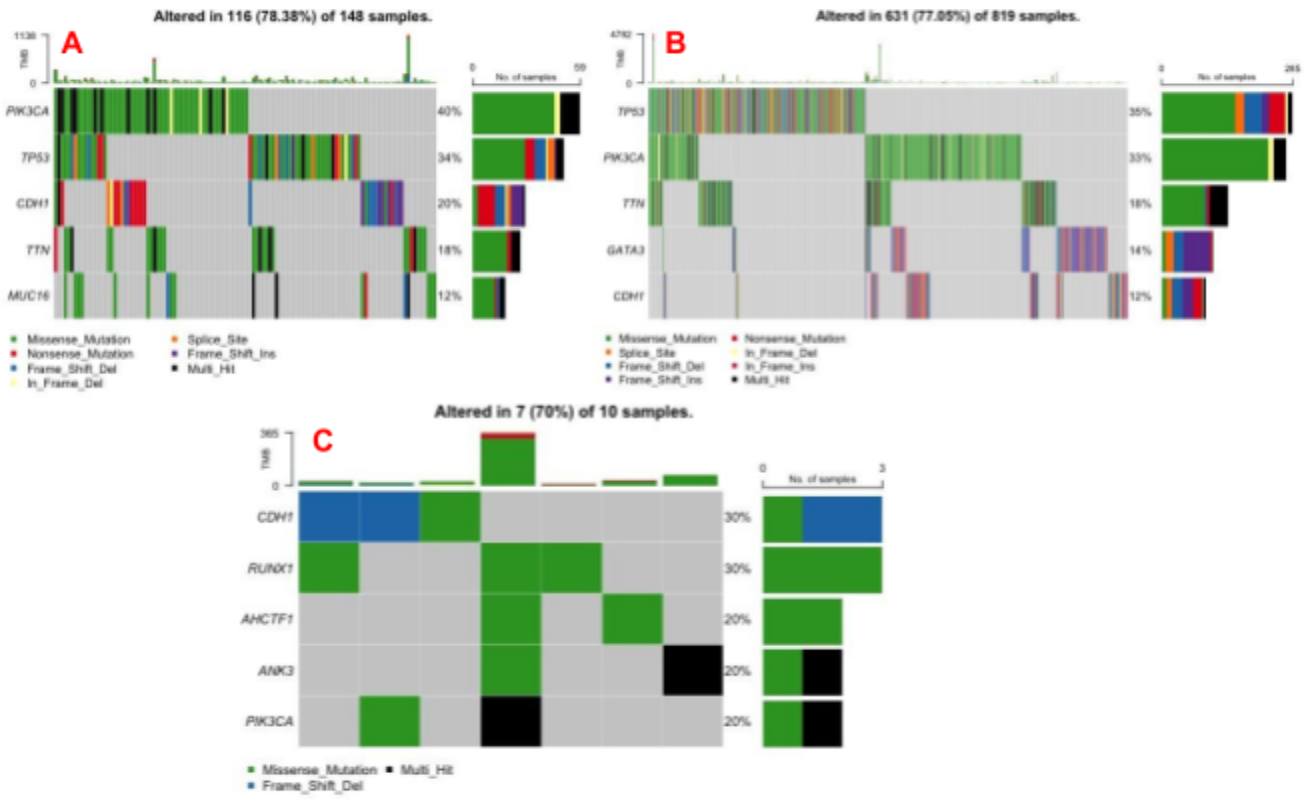

Figure 3. Oncoplots of the top 5 most common mutations for patients with a full recovery after radiation therapy (A), patients without radiation therapy (B), and patients without a full recovery after radiation therapy (C). The sample size of each group and the percentage of the sample that has the specific gene mutations are displayed at the top of each section.

In general, missense mutations are the most common type of mutations (Figures 3A, 3B, 3C). Fully recovered and non-radiation-treated patients had pretty similar top mutations–both groups had top mutations in the PIK3CA, TP53, TTN, and CDH1 genes with fully recovered patients having a unique top mutated gene of MUC16 and non-radiation-treated patients having a unique top mutated gene of GATA 3 (Figures 3A, 3B). However, the order in which the top mutated genes appear are in different order for the groups of patients that fully recovered and those who did not receive radiation therapy. Patients that did not fully recover have 4 unique top mutated genes of CDH1, RUNX1, AHCTF1, and ANK3 (Figure 3C). Patients that did not fully recover had a small sample size of 10 patients so the data collected is not completely reliable (Figure

3C). Finally, we wanted to examine the transcriptome of radiation patients and look at the over/underexpression of RNA of radiation patients.

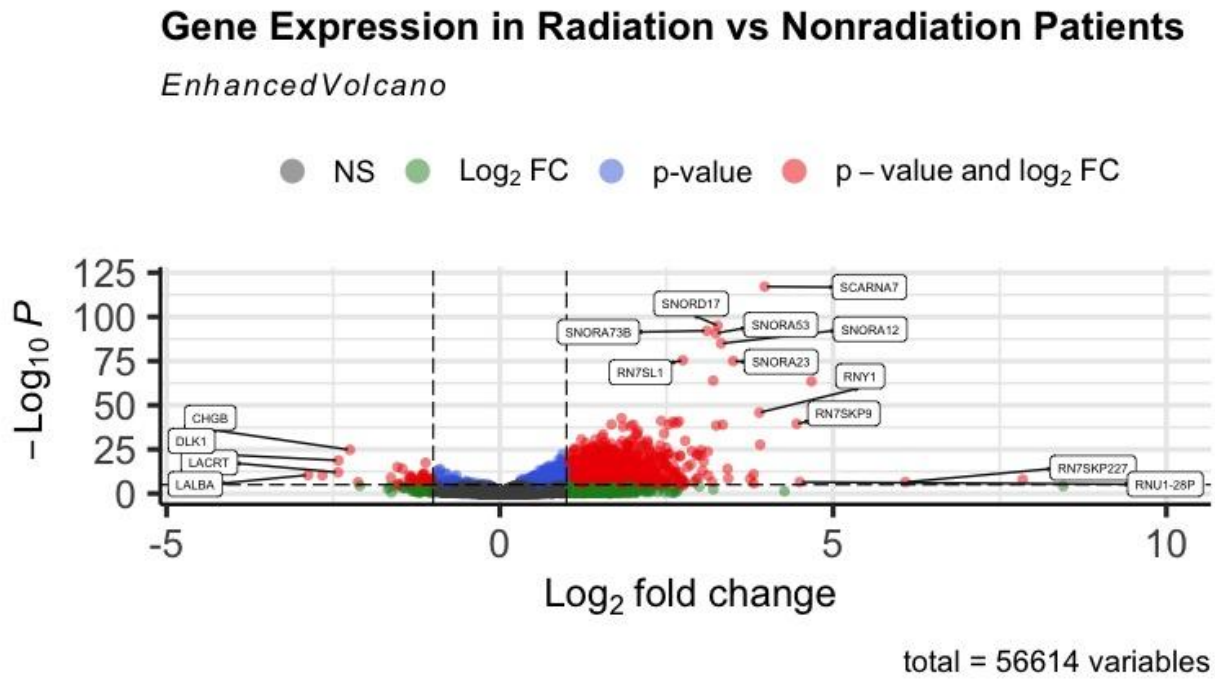

Figure 4. Volcano plot comparing RNA expression of radiation-treated patients as compared to non-radiation-treated patients as a baseline with adjustments based on pathologic stage and gender. Red values are those that had a significant enough p-value and a great enough change (either up or downregulated).

It is known that gender and the stage of cancer affects the transcriptome of cancer cells, so those were held constant while we compared the RNA expression of radiation patients in comparison to non-radiation patients (Huo et al.; Lopes-Ramos et al.) We found that the red genes on the right (eg. RNU1-28P, RNY1, SCARNA7, etc) are upregulated in radiation patients while the red genes on the left (eg. LACRT, LALBA, etc.) are significantly downregulated in radiation patients.

IV. Discussion

The age of patients and radiation therapy effectiveness we found was not strongly correlated we found (Figure 1). This is most likely due to the variable effects of cancer therapy and cancer progression and effects in older patients (Smith GL, Smith BD). Since cancer is a multifaceted disease, it is hard to predict the efficacy of the therapy based on age alone and requires a consideration of more factors. Looking at the survivorship curve in Figure 2, the data supports observations in Miller et al. reviewing cancer survivorship with a section on breast cancer. Typically, radiation therapy is only given to severe cases of cancer, however, when effective, provides a better prognosis than other types of therapy dealing with that stage of cancer (Miller et al.). Therefore, it makes sense that cancers that are not treated fully by radiation therapy are aggressive cancers that would result in a worse prognosis than the average breast cancer, supporting our evidence found in Figure 2. Looking at Figure 3A, we see that a unique highly mutated gene for radiation patients is MUC16. MUC16 is known to be highly overexpressed in cancer tissues and is known to promote tumorigenesis (Aithal et al.). Since radiation therapy works by causing DNA damage, damaging DNA that contains MUC16 would likely help reduce the tumorigenicity of a cancer cell and result in an effective treatment, showing the prevalence of MUC16 in patients that received radiation therapy and recovered. For patients that recieved radiation therapy but didn't fully recover, CDH1 is a possible biomarker of breast ductal cancer as in Adib et al. so its high presence in the cohort is supported by the current literature. However, a more interesting finding is the presence of mutations in the RUNX1 gene in Figure 3C. Mutations in RUNX1 have been found to be incredibly correlated with cancer malignancy and prognosis as it activates oncogenic signaling pathways, supporting our findings that patients that don't recover from radiation therapy likely have more aggressive forms of cancer (Tuo et al.). Finally, looking at the transcriptomic data, the overexpression of SNORA

genes is telling (Figure 4). Most of the patients in the radiation-treated patients fully recovered after radiation therapy, as there were more fully recovered patients (148) than not fully recovered patients (10) (Figures 3A, 3C). SNORA genes produce snoRNAs which modify eukaryotic RNA (Dieci et al.). A high prevalence of SNORA in patients could mean a higher modification of RNA to normal levels which could promote full recovery and better chances of survival as shown in Figure 3A.

It would be worth studying the extent of snoRNA control more to learn about what actual pathways they control and modify as there are a wide variety of SNORA genes that are known but not biochemically studied. This could better improve understanding of transcriptomic and epigenetic data and its effect on cancer. It would also be worth studying radiation patients that didn't fully recover from cancer as they most likely have the most aggressive forms of cancer that are very unique. A larger sample size than the one provided by TCGA would provide for a stronger test of mutated genes and would provide more insight into therapies for aggressive breast cancers. From our data, it can be hypothesized that RUNX1 and CDH1 mutations are likely a marker of aggressive breast cancers and can be tested for in order to administer stronger therapies than normal. This can be further investigated with a larger sample size for a stronger power of the test. It will also be worth looking at the actual factors that affect cancer therapy efficacy in tandem as looking at only one clinical factor is not hugely informative.

V. References

Adib, E. et al. CDH1 germline variants are enriched in patients with colorectal cancer, gastric cancer, and breast cancer. Br J Cancer 126, 797–803 (2022). https://doi.org/10.1038/s41416-021-01673-7

Aithal, A et al. MUC16 as a novel target for cancer therapy. Expert Opin Ther Targets. 2018 Aug;22(8):675-686. doi: 10.1080/14728222.2018.1498845. Epub 2018 Jul 26. PMID: 29999426; PMCID: PMC6300140.

Dieci, Giorgio et al. "Eukaryotic snoRNAs: a paradigm for gene expression flexibility." Genomics vol. 94,2 (2009): 83-8. doi:10.1016/j.ygeno.2009.05.002

Fialová, D., Laffon, B., Marinković, V. et al. Medication use in older patients and age-blind approach: narrative literature review (insufficient evidence on the efficacy and safety of drugs in older age, frequent use of PIMs and polypharmacy, and underuse of highly beneficial nonpharmacological strategies). Eur J Clin Pharmacol 75, 451–466 (2019). https://doi.org/10.1007/s00228-018-2603-5

Huo, T et al. Colorectal cancer stages transcriptome analysis. PLoS One. 2017 Nov 28;12(11):e0188697. doi: 10.1371/journal.pone.0188697. PMID: 29182684; PMCID: PMC5705125.

Lopes-Ramos, Camila M., et al. "Genome-Wide Sex and Gender Differences in Cancer." Frontiers in Oncology, vol. 10, 23 Nov. 2023, https://doi.org/https://doi.org/10.3389/fonc.2020.597788.

Łukasiewicz, Sergiusz et al. “Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review.” Cancers vol. 13,17 4287. 25 Aug. 2021, doi:10.3390/cancers13174287

Miller, Kimberly D., et al. “Cancer Treatment and Survivorship Statistics, 2022.” CA: A Cancer Journal for Clinicians, vol. 72, no. 5, 2022, pp. 409–436., https://doi.org/10.3322/caac.21731.

Smith GL, Smith BD. Radiation treatment in older patients: a framework for clinical decision making. J Clin Oncol. 2014 Aug 20;32(24):2669-78. doi: 10.1200/JCO.2014.55.1168. Epub 2014 Jul 28. PMID: 25071132; PMCID: PMC4876341.

Tuo Z et al. RUNX1 is a promising prognostic biomarker and related to immune infiltrates of cancer-associated fibroblasts in human cancers. BMC Cancer. 2022 May 9;22(1):523. doi: 10.1186/s12885-022-09632-y. PMID: 35534796; PMCID: PMC9088136.