# A Multi-Omic Regressive Analysis of Hormonal Receptor Status in Breast Cancer

## INTRODUCTION

Breast cancer ranks among the most fatal forms of cancer, and it is the most frequent malignancy among women (Harbeck, N. et al., 2019). Both the fatality rates and incidence of BC have increased over the past few decades, and there are an estimated 2.3 million new cases of breast cancer every year (Łukasiewicz, S. et al., 2021). There are a multitude of well-evidenced risk factors for BC, including demographic, reproductive, hormonal, hereditary, breast-contextual, and lifestyle-related factors affecting the prevalence and frequency of this indication (Momenimovahed, Z. et al., 2019) . While patients with early-stage, non-metastatic BC are successfully cured in nearly 80% of cases, more advanced breast cancer involving organ metastases is typically deemed incurable with currently available interventions (Harbeck, N. et al.). As a result, it is in the best interests of physicians and researchers alike to seek insights that fuel earlier detections of BC prior to the metastatic stages.

BC possesses a complex histology, as it can be allocated into a number of histological types with varying characterization (Thennavan, A. et al., 2021). The variances within cancerous histology entail structural modifications within discrete processing stages, from the genome to proteome. A corollary of these conditions is the manifestation of diverse cancer phenotypes, stipulating that the research into the underlying mechanisms of any given form of cancer must be analyzed within a number of different omic layers, namely genomics, transcriptomics, and

proteomics (Das, T. et al., 2020). These efforts have been facilitated by The Cancer Genome Atlas (TCGA), a publicly financed effort that employs extensive genome sequencing and integrated multi-dimensional omics data analyses to catalog and identify potentially cancer-causing genomic changes. In making publicly accessible cancer genetic databases possible for over 30 human tumors, TCGA enables the development of diagnostic procedures, medical care standards, and ultimately cancer prevention (Tomczak, K. et al., 2015).

Hormonal receptors have been consistently linked to specialized forms of breast cancer, namely progesterone receptors (PR) and estrogen receptors (ER). Prior studies have alluded to differential outcomes based on receptor positivity, as positive receptor status has been demonstrably associated with favorable predictive features such as diminished cell proliferation and burgeoning evidence of histological tumor differentiation (Osborne, C. K. , 1998). Furthermore, the vast majority of BC-related deaths in the US and Europe are caused by Hormone receptor (HR)+ breast cancer (Buqué, A et al., 2020). However, little is known about how these receptors operate independently of one another, particularly in regards to the manifestation of their genetic and molecular epidemiology. To address these deficiencies within the field, we determined that The Cancer Genome Atlas' Breast Carcinoma database (TCGA-BRCA) had unique utility in appraising the differential effects of PR and ER in patients with breast cancer.

This study utilized the clinical and mutation-annotated format (MAF) data within the TCGA-BRCA multi-omic database to drive our primary object, namely an inquiry into differential survivability of hormone receptors, both locally (positive vs negative within a subtype of hormonal receptors) and generally across the two presented types of hormonal

receptors. Clinically-annotated oncoplots and co-oncoplots were devised to determine the most significant effects of receptor selectivity upon gene expression for the hormonal receptors, and the genes with the highest degree of differentiation between the positive and negative subtypes of hormones were then analyzed in co-lollipop plots to spatially contextualize the underlying stimulus for differentiation. The clinically-inclined segments of the database were utilized in a survivability analysis for each of the two hormonal receptors as seen by Kaplan-Meier plots. Secondary outcomes of this study included a comparison of the hormonal receptors to surmise more diagnostic aspects of breast cancer. A bar chart was created to compare the relative frequencies of both hormonal receptors relative to one another, and violin plots were used to ascertain the effect of each hormonal receptor on the signal result range within the Fluorescence *in-situ* Hybridization (FISH) diagnostic procedure.

## METHODS

Analysis in the study was primarily derived upon breast cancer multi-omics data, which was accessed from The Cancer Genome Atlas (TCGA) with the ascension code 'TCGA-BRCA'. A number of packages central to this analysis were installed and loaded, including TCGAbiolinks, maftools, survminer, ggplot2, and vioplot. The two datasets central to this search included clinical and mutation annotation format (MAF) data; each was independently queries and prepared using GDCquery and GDCprepare, respectively. Upon thorough examination of the variables measured within the dataframe and a subsequent thematic analysis to deduce the most clinically relevant issues within breast cancer research, hormonal receptor positivity was determined to be the category of interest for this study.

Upon completion of the clinical dataframe, boolean masking was employed to create two columns of filtered data for the hormone receptors within the dataset, namely the status of progesterone and estrogen receptors. The two columns, named PR_category and ER_category, assigned labels of specified positivity based on the already existing receptor status columns in the dataframe. These updates to the dataframe were processed via the maftools package, which processed the maf data through the 'read.maf' function. An initial oncoplot was developed with the oncoplot function and processed maf data, involving both the subsetted columns as clinical features, to plot the differential gene expression between progesterone and estrogen receptors. The dataframe was then subsetted into two separate data frames for each of the hormonal receptors, and boolean masking was used to store and subset the patient barcodes. The co-oncoplot function was then used to draw two oncoplots side by side, providing a differential analysis of gene expression within each of the two hormonal receptors. The three genes with the most disparate frequencies across receptors were highlighted by co-lollipop plots, which also utilized the subsetted binary objects for both receptors within the lollipopPlot2() function. Kaplan-Meier curves were drawn via the survival and survminer packages to depict the differential effects of the two receptor statuses on survival, ultimately yielding a measure of statistical significance. Survival time was dependent on available data within the dataframe; if the "days to death" variable was not marked as NA, then the value was used to indicate survival time, and if it was marked as NA, then the "days to last followup" variable was used in its place. This brought a further means of abridged contextualization to the distribution of receptor selectivity on patient outcomes.

Additionally, violin plots were developed through the vioplot package to establish new evaluative paradigms within the differential effects of both hormonal receptors, specifically within the Fluorescence in-situ Hybridization (FISH) diagnostic procedure of determining the signal result range of Chromosome 17. FISH analysis is focused on the detection of HER2 amplification, which has been shown to be associated with Chromosome 17 polysomy; such polysomy has been demonstrated to predict metastatic spread in breast cancer (Jiang, H. et al., 2014). Since FISH analysis is effectively used as a proxy to predict the malignant potential of breast cancer, the violin plot analysis investigated the differential outcomes of hormonal receptors upon the range of this FISH diagnostic procedure. Furthermore, a relative frequency bar chart was created with the aplpack and ggplot2 packages to compare each of the pairwise receptor frequencies (ie ER-negative with PR-negative, ER-positive with PR-negative, etc.), thereby corroborating previous experimental work and bolstering the perceived representativeness of the TCGA-BRCA database.

## RESULTS

### *I. Primary Outcomes*

Our studies indicate the utility of independently appraising progesterone receptor (PR) and estrogen receptor (ER) status upon survivability. We found that BRCA patient populations demonstrated a statistically significant difference ($p = .0046$) in survivability probability between the PR-positive and PR-negative patients. (Figure 1). A similar, albeit more statistically muted finding was determined for estrogen receptor status, where ER-positive patients demonstrated a significantly lower survival rate ($p = 0.025$) than ER-negative patients (Figure 2).

Oncoplot analysis indicates heterogenous patterns of gene expression across hormonal receptor status. The pattern of annotated mutations for both receptors was largely similar, implying similar thresholds of differential mutations within BRCA patients regardless of receptor status (Figure 3). Co-oncoplots in subsetted data for progesterone and estrogen receptors indicated genes with the highest rates of mutation, with the differential mutation analysis across positive and negative subtypes also indicated. Not only were the top 3 most mutated genes common across PR and ER subsets, but the differential mutation frequencies were also largely proportional (Figure 3, 4).

The top 3 genes with the greatest mutation frequency and differences within said mutation frequencies were plotted by co-lollipop analysis between each type of receptor. The gene *PIK3CA* exhibited limited variation in the mutation binding sites or type for each of the PR and ER subsets, as the spatial gene expression distribution was similar for both positive and negative receptor subtypes (Figure 6, 7). This finding was consistent with that of the gene *TP53*, with limited differences in the spatial distribution of gene expression across the gene motifs and domains (Figure 10, 11). However, there appeared to be variations in mutation spatial distribution in the gene *CDH1* for both ER and PR subgroups. *CDH1* mutation expression was spatially skewed left for PR-negative patients, while it was much more evenly distributed for PR-positive patients (Figure 8). There was similarly a greater diversity and frequency of mutation types for PR-positive patients as compared to PR-negative patients. These variations were even more pronounced in estrogen receptor subgroups, where there were significant spatial differences in the distribution of mutations between ER-positive and ER-negative patients (Figure 9).

*II. Secondary Outcomes*

The secondary outcomes to evaluate the diagnostic utility of hormonal receptor data pitted the relative effects of each receptor selectivity against one another. The relative binary frequencies of PR and ER statuses indicated the positive correlation between PR and ER positivity, as the majority of receptor-receptor pairings are commensurable (Table 1). Violin plots surmised the relative effect of receptor presence on Fluorescence *in-situ* Hybridization (FISH) result range, exhibiting an apparent association between receptor positivity and result range. PR positive patients exhibit a slightly greater range of result outcomes compared to PR negative patients; while the density shape and lower thresholds are similar, the PR positive patients demonstrate much higher extreme thresholds, including 75th percentile distribution and outliers (Figure 14). ER positive patients exhibit a much greater signal result range compared to ER negative patients; the upper and lower percentile thresholds, as well as the general density shape and outlier distribution are far more pronounced within the ER positive patients (Figure 15).

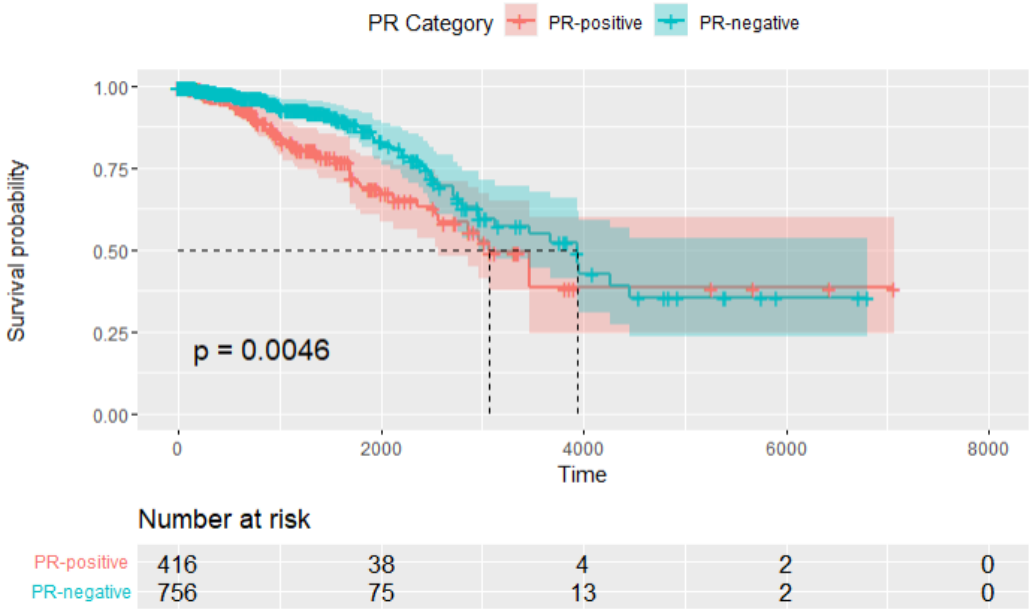



**Figure 1: Kaplan-Meier analysis shows that patients demonstrating progesterone receptor (PR) positivity have a significantly lower survival rate ($p$ = 0.0046) than patients with negative PR status.** Plots were made using the survplot function on R to plot the survival probabilities of various anatomical regions, using breast cancer clinical data retrieved from TCGA-BRCA. Time is calculated here as the time from diagnosis to the death or final follow up, depending on data availability.

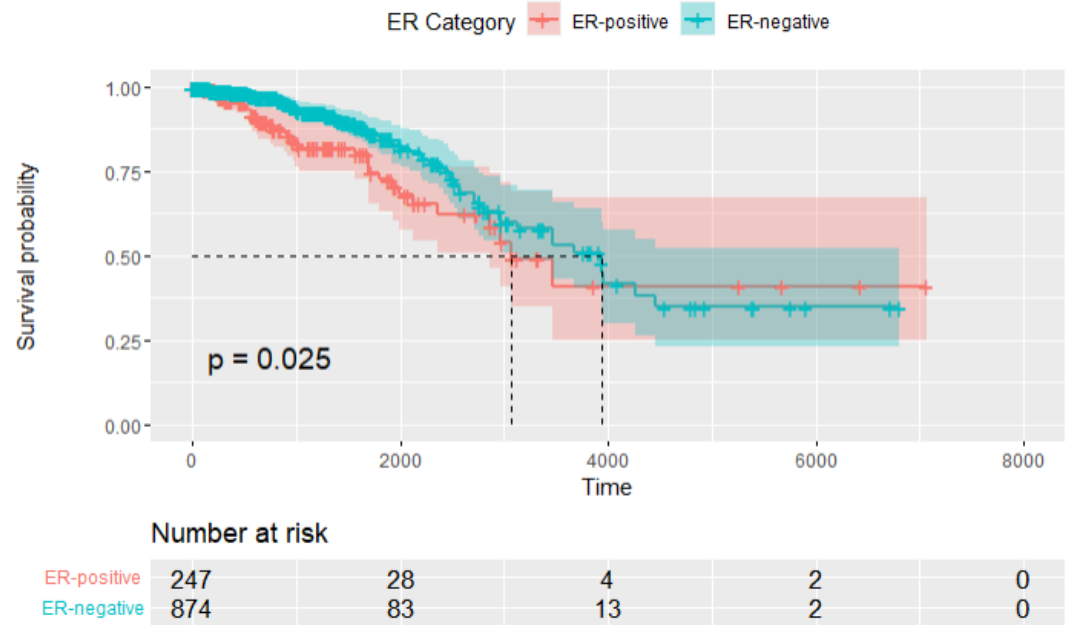


**Figure 2: Kaplan-Meier analysis shows that patients demonstrating estrogen receptor (ER) positivity have a significantly lower survival rate ($p$ = 0.025) than patients with negative ER status.** Plots were made using the survplot function on R to plot the survival probabilities of various anatomical regions, using breast cancer clinical data retrieved from TCGA-BRCA. Time is calculated here as the time from diagnosis to the death or final followup, depending on data availability.

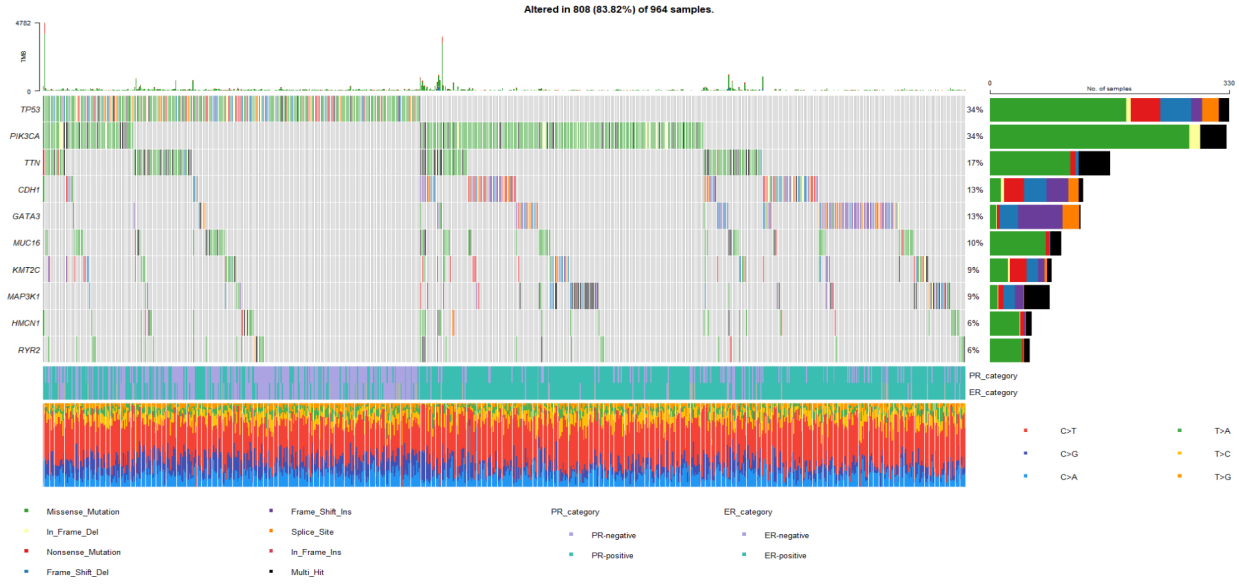



**Figure 3: Oncoplot suggests similar thresholds of differential mutations within BRCA patients in the mutation annotation format (MAF) data.** Within the annotation tab, the annotation colors for both hormonal receptor statuses seem to be largely common. Oncoplots were made using the R-package maftools showing total gene mutations and characterizing each gene by types of mutations. Mutation and clinical data were retrieved from TCGA-BRCA, with the data referencing nucleotide frequency, annotation frequency, and gene expression in breast cancer.

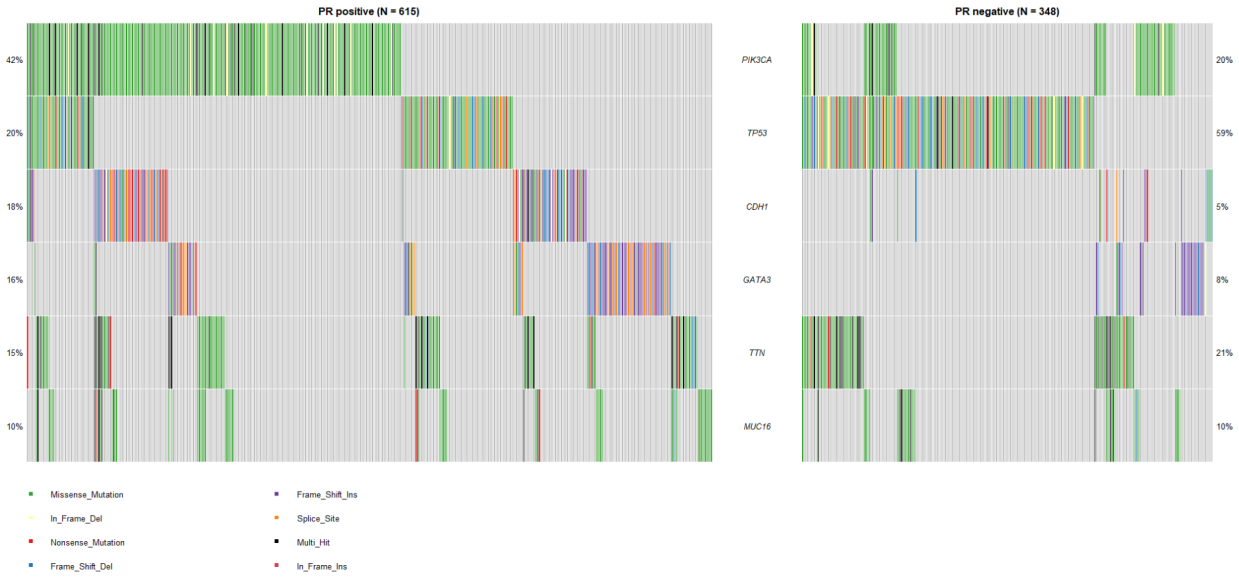



**Figure 4: Oncoplot analysis shows the top 6 genes with the highest rate of somatic mutations that were shared between patient groups with positive PR status and negative PR status.** Oncoplots were made using the R-package maftools showing total gene mutations and characterizing each gene by types of mutations. Mutation and clinical data were retrieved from TCGA-BRCA, with the data referencing gene expression in breast cancer.

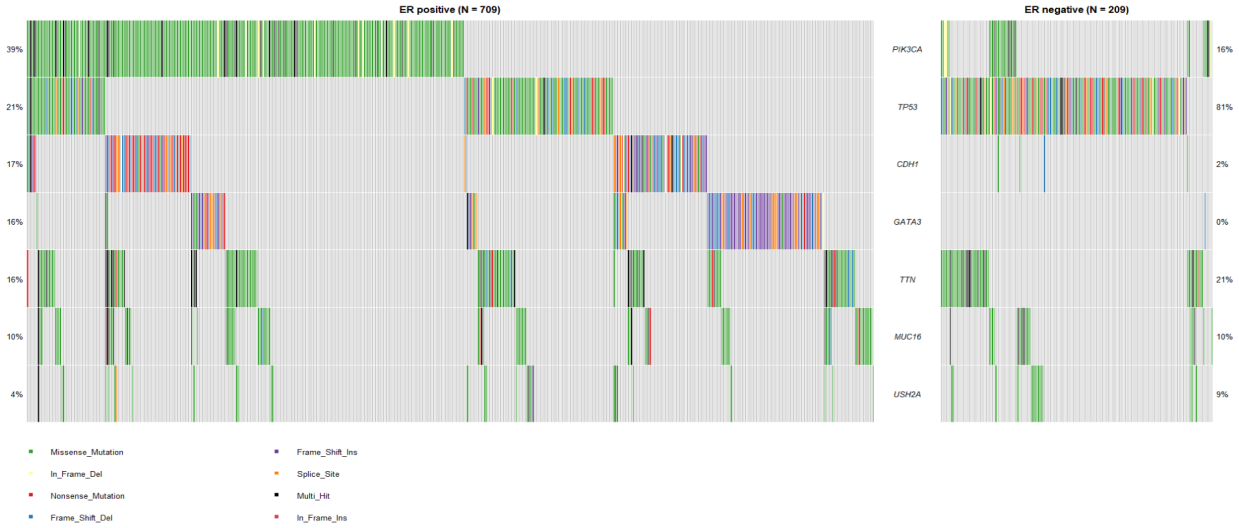



**Figure 5: Oncoplot analysis shows the top 6 genes with the highest rate of somatic mutations that were shared between patient groups with positive ER status and negative ER status.** Oncoplots were made using the

R-package maftools showing total gene mutations and characterizing each gene by types of mutations. Mutation and clinical data were retrieved from TCGA-BRCA, with the data referencing gene expression in breast cancer.

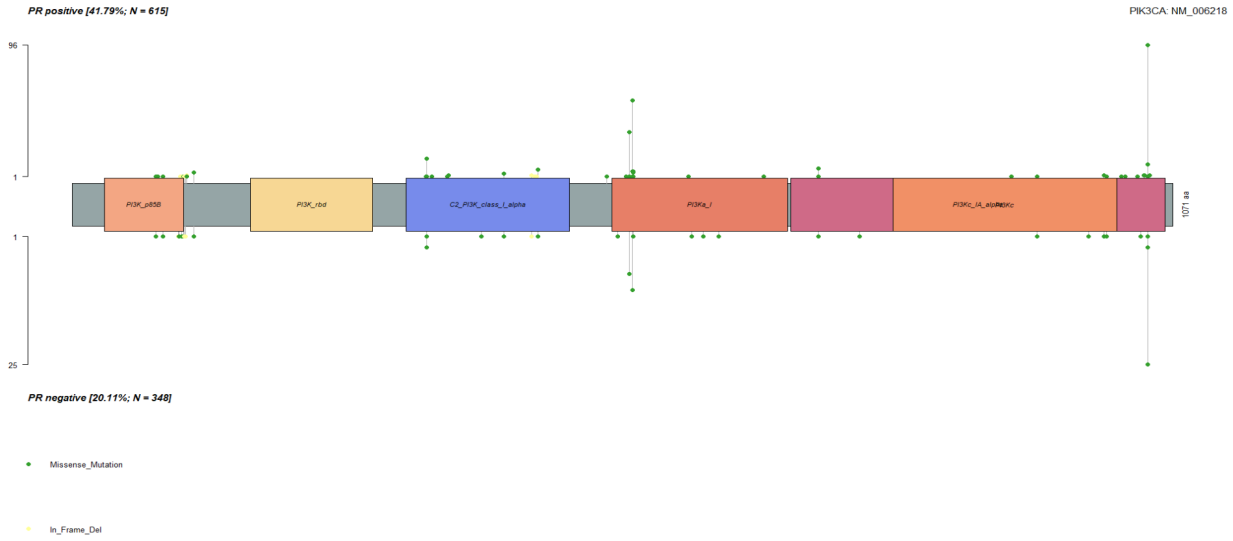



**Figure 6: Despite differences in mutation frequency, Co-lollipop analysis of PIK3CA between patients exhibiting PR positive and PR negative statuses reflects little variation across spatial regional sites or specific mutation types.** Co-lollipop plots were made using the function lollipopPlot2, retrieving spatial gene expression data among TCGA-BRCA patients.

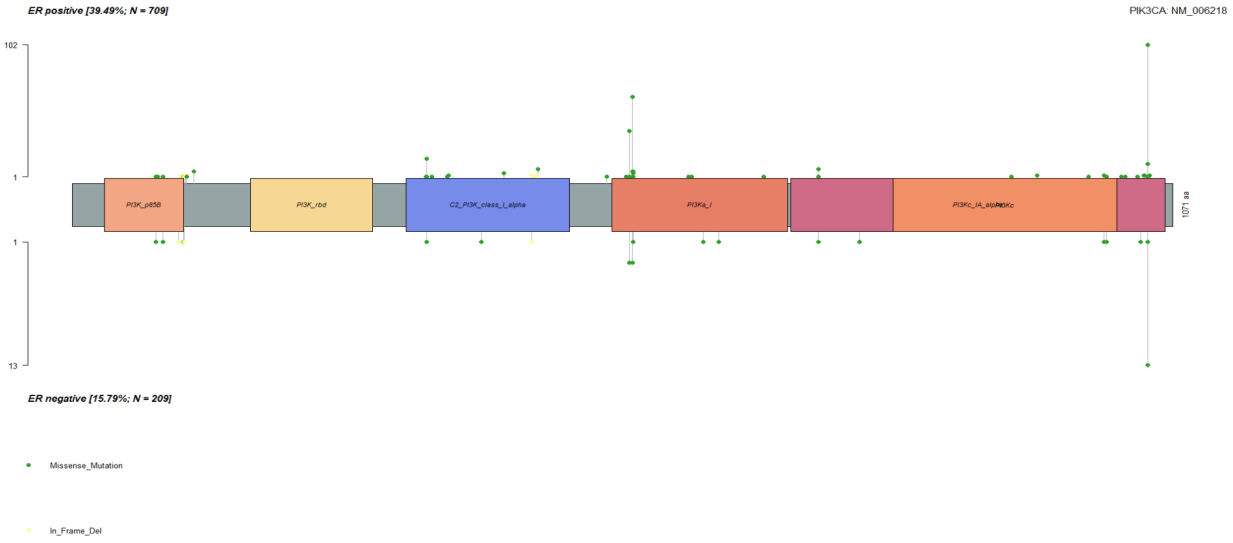


**Figure 7: Despite differences in mutation frequency, Co-lollipop analysis of PIK3CA between patients exhibiting ER positive and ER negative statuses reflects little variation across spatial regional sites or specific mutation types.** Co-lollipop plots were made using the function lollipopPlot2, retrieving spatial gene expression data among TCGA-BRCA patients.

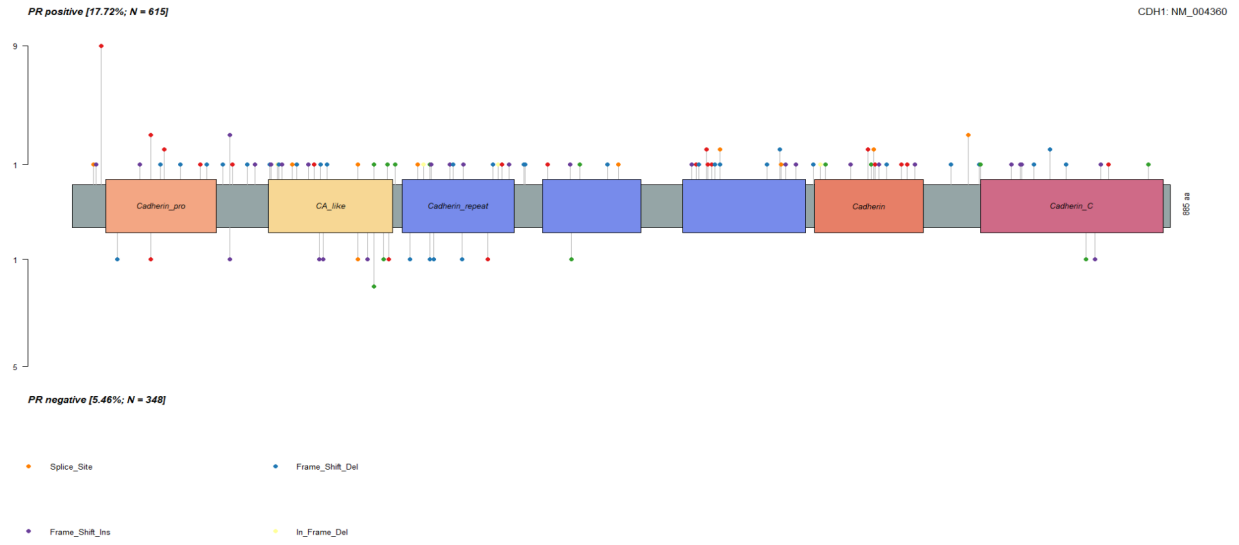



**Figure 8: Amidst differences in mutation frequency, Co-lollipop analysis of CDH1 between patients exhibiting PR positive and PR negative statuses reflects significant variation across spatial regional sites or specific mutation types.** PR negative status appears to have significantly fewer mutations, especially on the right side of the gene domains and motifs depicted. Co-lollipop plots were made using the function lollipopPlot2, retrieving spatial gene expression data among TCGA-BRCA patients.

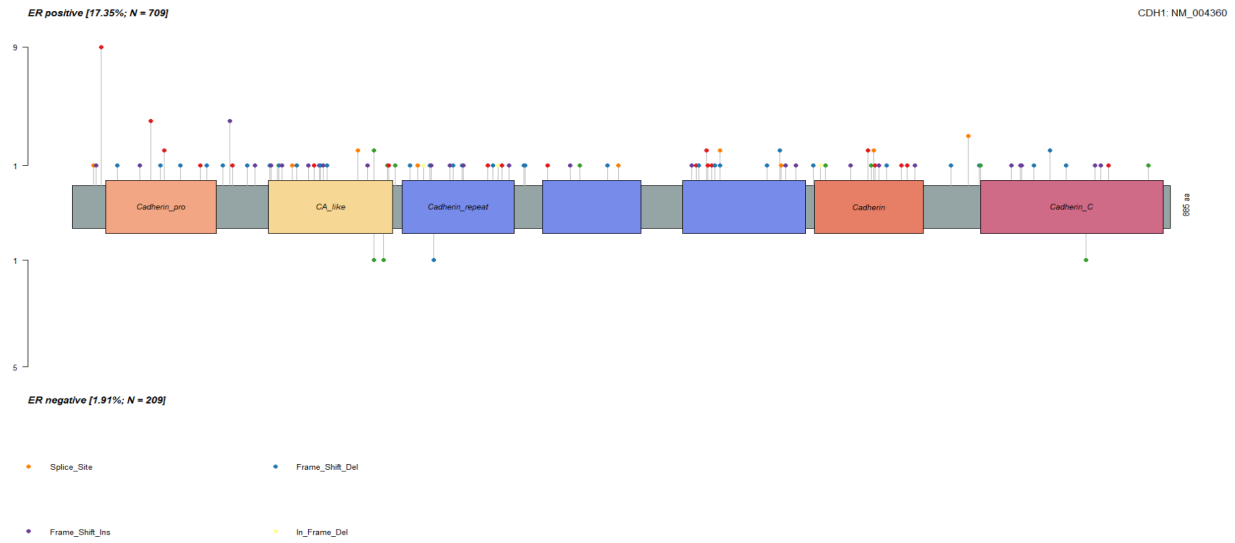



**Figure 9: Amidst differences in mutation frequency, Co-lollipop analysis of CDH1 between patients exhibiting ER positive and ER negative statuses reflects significant variation across spatial regional sites or specific mutation types.** ER negative status appears to have significantly fewer mutations, with a minimal distribution across the gene domains and motifs depicted. Co-lollipop plots were made using the function lollipopPlot2, retrieving spatial gene expression data among TCGA-BRCA patients.

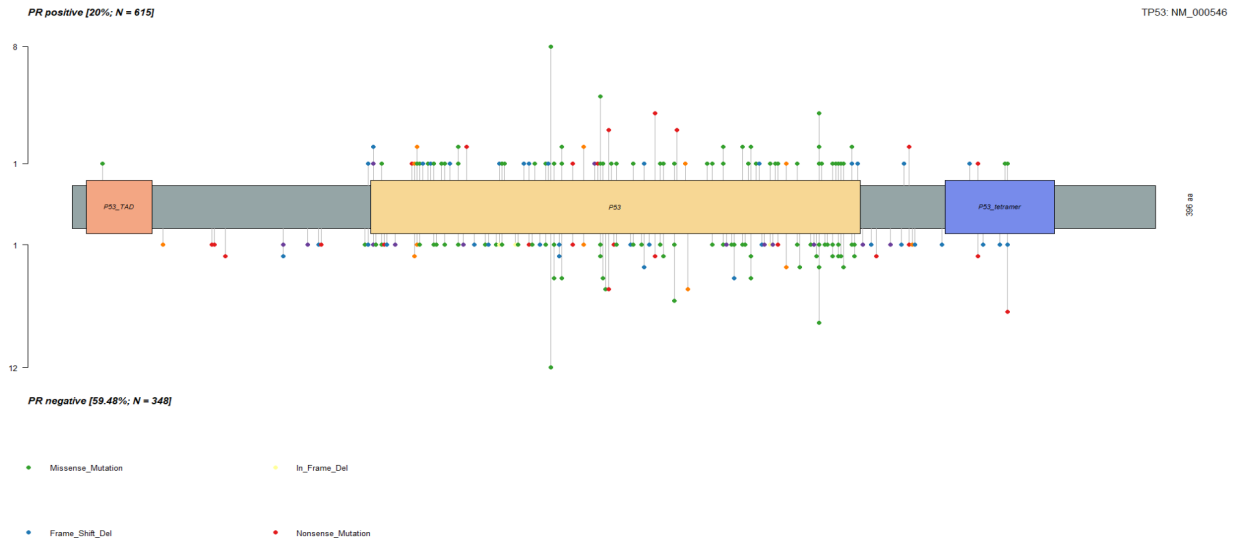


**Figure 10: Despite differences in mutation frequency, Co-lollipop analysis of TP53 between patients exhibiting PR positive and PR negative statuses reflects little variation across spatial regional sites or specific mutation types.** Co-lollipop plots were made using the function lollipopPlot2, retrieving spatial gene expression data among TCGA-BRCA patients.

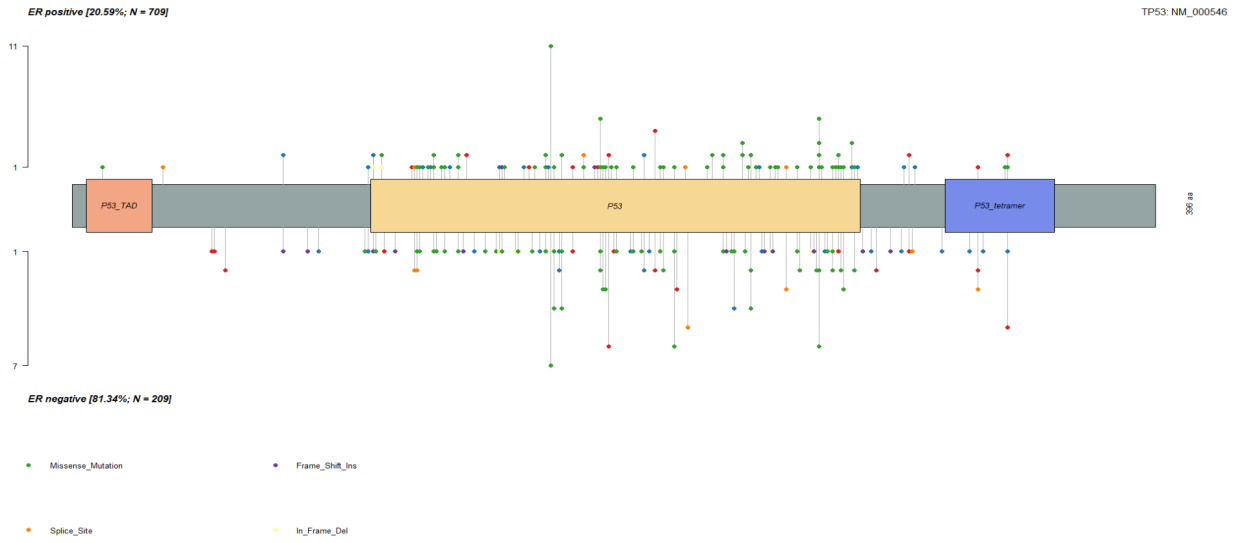
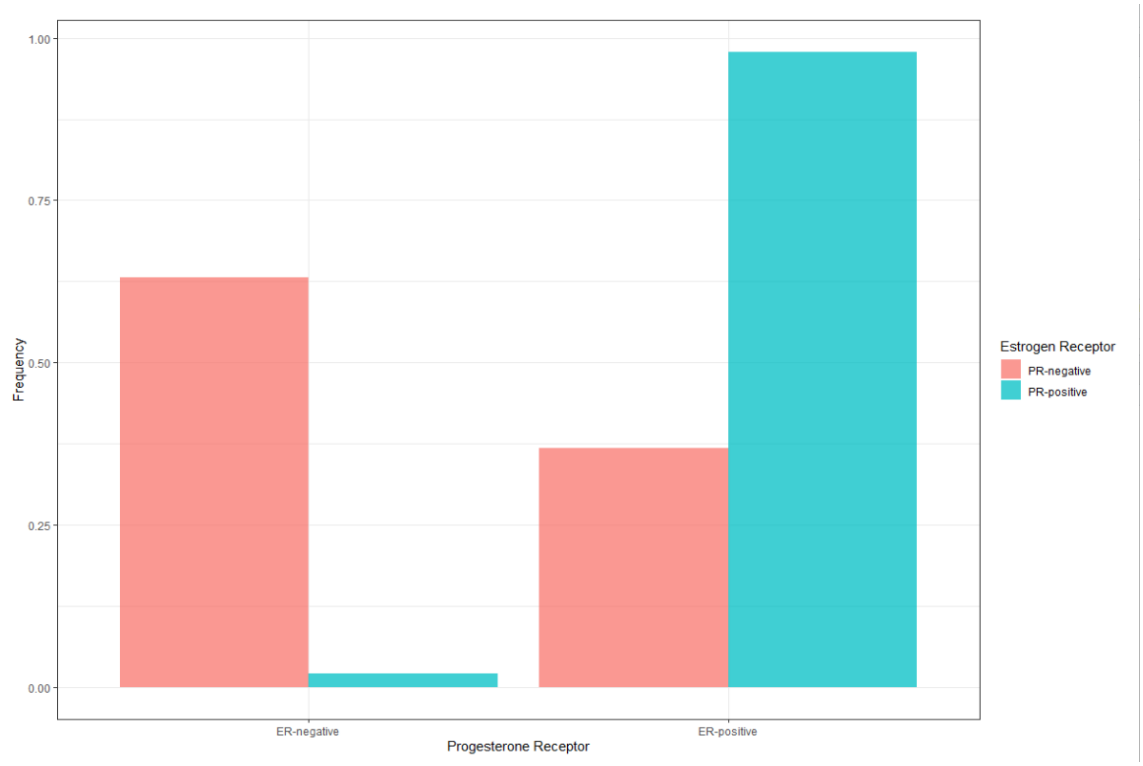



**Figure 11: Despite differences in mutation frequency, Co-lollipop analysis of TP53 between patients exhibiting ER positive and ER negative statuses reflects little variation across spatial regional sites or specific mutation types.** Co-lollipop plots were made using the function lollipopPlot2, retrieving spatial gene expression data among TCGA-BRCA patients.

**Table 1: Relative binary frequencies of progesterone and estrogen receptor status**



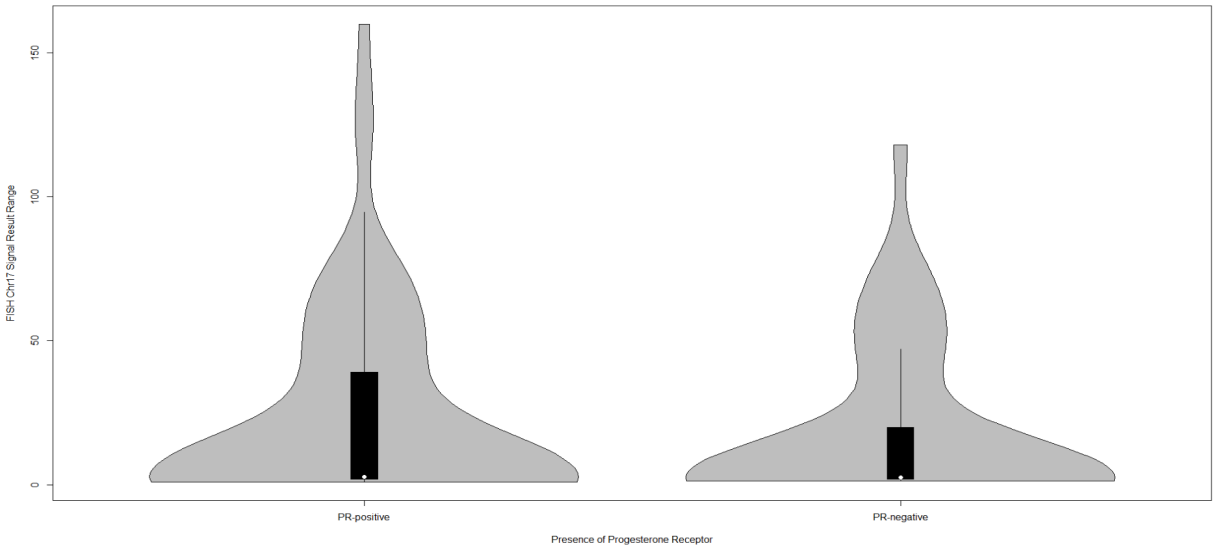



**Figure 14: Violin plot analysis shows that patients demonstrating PR positivity appear to exhibit a somewhat larger Fluorescence *in-situ* Hybridization (FISH) Chromosome 17 Signal Result Range.** PR-positive patients reflect slightly greater ranges of outcomes, especially in regards to outliers. The density shape and lower thresholds seem to follow similar distribution. Plots were made using the vioplot function on R to plot the density distributions across the signal result range for each receptor type, using breast cancer clinical data retrieved from TCGA-BRCA.

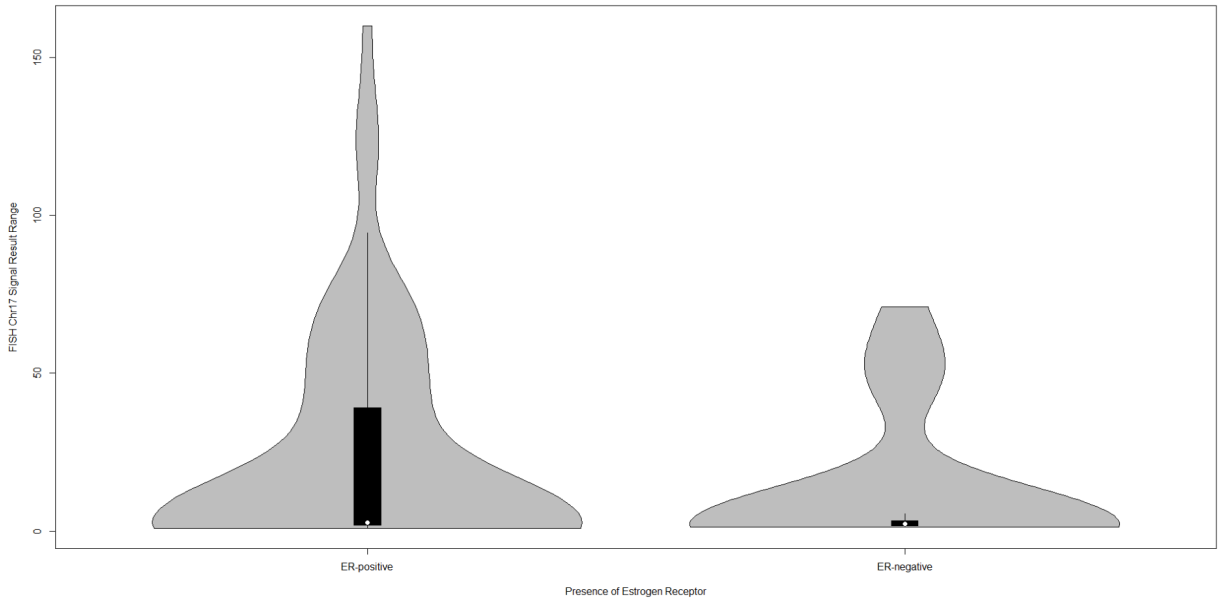



**Figure 15: Violin plot analysis shows that patients demonstrating estrogen receptor (ER) positivity appear to exhibit a significantly larger Fluorescence *in-situ* Hybridization (FISH) Chromosome 17 Signal Result Range.** ER-positive patients reflect far greater ranges of outcomes, in regards to the upper percentile thresholds and outliers. Plots were made using the vioplot function on R to plot the density distributions across the signal result range for each receptor type, using breast cancer clinical data retrieved from TCGA-BRCA.

## DISCUSSION

Given the diversity of treatment options and prevalence of breast cancer there are many diverging perspectives on the treatment of this indication. A collection of high quality multi-omics data is critical to establishing more context and insights into the effectiveness of current interventions and the potential application of research-based insights to develop future interventions. Such insights are intuitively critical for the effective implementation of these interventions to meet both physician and patient needs, ultimately providing utility towards cancer prevention.

This paper evaluates the differential effects of the predominant hormonal receptors present in breast cancer, progesterone receptors and estrogen receptors, with the primary objective of assessing the independent effects of these hormonal receptors on survivability and gene expression. A secondary objective of this study was to appraise the potential diagnostic ability of the hormonal receptor data. To our knowledge, no review has specifically examined breast cancer to this degree of comprehensiveness or scope. Our results indicate that there is a statistically significant effect of hormonal receptor positivity on survivability within breast cancer patients (Figure 1, 2). This corroborates the findings of Buque et al, which reported that positive hormonal receptor statuses were the predominant cause of death. While this may seem intuitive, a critical aspect of the scientific process entails the probing of somewhat intuitive factors within a singular phenomenon to appraise whether those factors (in this case the specific positivities within each hormonal receptor type) have a differential effect on said phenomenon. Consequently, we can surmise that diagnostic and intervention implementation among patients

with a singular positive receptor should still mandate a requisite level of urgency, as our findings indicate the independent effects of a singular positive receptor upon survivability. Even though positive-negative pairwise receptor pairs are somewhat rare (Table 1), our findings can address the more sporadic cases of breast cancer that aren't as intently focused on. Even with their minimal population proportions, these cases still comprise thousands of cases considering the widening range of outcomes for breast cancer.

Rates of mutation across genes in both PR and ER subgroups followed similar patterns (Figure 3, 4, 5). The most common mutations between PR subgroups and ER subgroups were identical, indicating that the type of receptor (PR vs ER) has a negligible effect on mutation expression. However, both PR and ER negative patients exhibited much less marked rates of mutation among these genes. Similarly, when plotting the spatial distribution of these genes across each of the receptor subgroups, we expected to find similar differences as seen by the co-oncoplots. While *PIK3CA* and *TP53* did not deviate significantly between positive and negative receptor status, we surprisingly found a significant difference in mutation spatial distribution between positivity statuses for the *CDH1* gene (Figure 6-11). This is a novel finding, implying that CDH1 mutation is correlated with positive receptor statuses for PR and ER subtypes. The bigger takeaway, however, is that *CDH1* has potential viability as a biomarker for breast cancer. Rates of mutation within this gene could be monitored, and the diagnostic utility of this gene can be realized to not only predict the metastatic potential of the cancer, but could even identify individuals at risk of breast cancer before they test positive for cancer. The differential effects of mutation spatial distribution for *CDH1* needs to be studied in greater detail to determine whether it holds unique efficacy in predicting and preventing breast cancer.

Finally, the violin plots were of particular interest, as they indicated that receptor positivity was associated with a greater signal result range of Chromosome 17 as ascertained by the FISH diagnostic measure (Figure 14, 15). This was a surprising finding, as FISH analysis is not inherently meant to measure hormonal receptor status. On the contrary, FISH analysis evaluates the amplification of HER2 receptor, an epidermal growth factor that operates as an indicator for metastatic spread in breast cancer (Jiang et al., 2018). This indicates that hormonal receptor positivity is correlated with the amplification of HER2 receptor, thereby implicating its independent fidelity in projecting metastatic potential in breast cancer. Jiang alluded to the importance of hormonal receptor positivity in breast cancer detection, but there was no formal analysis beyond such conjecture (Jiang et al., 2018). Specifically in our data, we found that the differential effect of PR positivity on signal result range was much greater than that of ER positivity, despite both receptors' positivity status clearly outpacing their respective negative receptor statuses. Thus, we can tentatively contend that ER positivity has more predictive power in the detection of breast cancer than PR positivity, although this contention will need to be corroborated by future studies. This is an important study, as this provides a realistic means of alerting patients about their risk for breast cancer before they are formally diagnosed. Furthermore, we can combine this finding with our previous findings to form a more holistic diagnostic and prognostic recommendation for patients who exhibit positivity within the two hormonal receptors discussed at length here.

While the implications of our research has been disseminated in a stochastic fashion across this section, we believe that the future direction of multi-omics breast cancer data is immensely tantalizing. Greater research into the effect of *CDH1* mutation distribution across

breast cancer prediction and prevention, as well as the potential of ER positivity in diagnostically assessing breast cancer, are the findings of this study with the greatest potential. While the goal of this analysis is indubitably the prevention of breast cancer, we hope that our study can fit a niche within this field by contributing to this fundamental resolution.

# Works Cited

Buqué, A., Bloy, N., Perez-Lanzón, M., Iribarren, K., Humeau, J., Pol, J. G., Levesque, S., Mondragon, L., Yamazaki, T., Sato, A., Aranda, F., Durand, S., Boissonnas, A., Fucikova, J., Senovilla, L., Enot, D., Hensler, M., Kremer, M., Stoll, G., … Galluzzi, L. (2020). Immunoprophylactic and immunotherapeutic control of hormone receptor-positive breast cancer. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-17644-0

Das, T., Andrieux, G., Ahmed, M., & Chakraborty, S. (2020). Integration of online OMICS-Data Resources for Cancer Research. Frontiers in Genetics, 11. https://doi.org/10.3389/fgene.2020.578345

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., & Cardoso, F. (2019). Breast cancer. Nature Reviews Disease Primers, 5(1). https://doi.org/10.1038/s41572-019-0111-2

Jiang, H., Bai, X., Meng, F., Zhang, C., & Zhang, X. (2014). Evaluation of chromosome 17 polysomy in breast cancer by fish analysis of whole nuclei, and its clinicopathological significance. Oncology Letters, 7(6), 1954–1958. https://doi.org/10.3892/ol.2014.2001

Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer: Targets and Therapy, Volume 11, 151–164. https://doi.org/10.2147/bctt.s176070

Osborne, C. K. (1998). Steroid hormone receptors in breast cancer management. Breast Cancer Research and Treatment, 51(3), 227–238. https://doi.org/10.1023/a:1006132427948

Thennavan, A., Beca, F., Xia, Y., Garcia-Recio, S., Allison, K., Collins, L. C., Tse, G. M., Chen, Y.-Y., Schnitt, S. J., Hoadley, K. A., Beck, A., & Perou, C. M. (2021). Molecular analysis of TCGA breast cancer histologic types. Cell Genomics, 1(3), 100067. https://doi.org/10.1016/j.xgen.2021.100067

Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). Review the cancer genome atlas (TCGA): An immeasurable source of knowledge. Współczesna Onkologia, 1A, 68–77. https://doi.org/10.5114/wo.2014.47136

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. Cancers, 13(17), 4287. https://doi.org/10.3390/cancers13174287