Name: Alan Xu

QBIO 490: Directed Research

Midterm Project

3/13/2023

Introduction

In 1990, cancer became the 2nd leading cause of death worldwide behind cardiovascular diseases (Fitzmaurice et al., 2015). In 2013 alone, 8.2 million patients succumbed to some form of cancer (Fitzmaurice et al., 2015). Among various forms of cancer, breast cancer is the most prevalent form of cancer within females (Łukasiewicz et al., 2021). Breast cancer is now recognized as the 5th leading cause of death in the population worldwide and in 2020 alone, 2.3 million people were newly diagnosed with breast cancer (Łukasiewicz et al., 2021). Clearly, breast cancer is a pressing public health issue that requires immediate attention from the scientific community.

Breast cancer is known to have a genetic component during its development. The most well-researched mutations include BRCA1 and BRCA2 mutations that are implicated in the DNA repair pathway (Łukasiewicz et al., 2021), other genes that are highly associated with breast cancer risk include TP53, CDH1, PTEN and STK11(Łukasiewicz et al., 2021). Around 15 – 20% breast cancers overexpress HER2, a membrane tyrosine kinase implicated in cancer cell proliferation (Krishnamurti & Silverman, 2014). There is no known molecular ligand for the HER2 receptor (Swain et al., 2023). Historically, $HER2^+$ breast cancer was regarded as one of the most aggressive types of breast cancers (Swain et al., 2023). However, discovery of antibody treatments against $HER2^+$ cancer cells, such as trastuzumab and more, has significantly improved the survivability of $HER2^+$ breast cancer (Swain et al., 2023). Under current standards

of care, survival rate of early $HER2^{+}$ breast cancer exceeds 90% (Swain et al., 2023), prompting HER2 as a potent molecular target for future cancer treatment.

Despite HER2-focused onco-therapies' success, $HER2^{+}$ cancer cells do eventually develop resistance to HER2-targetting therapies. Thus, this paper focuses on multi-omic analysis between $HER2^{high/low}$ breast cancer patients to discover if there are any molecular clues that separate the two patient populations. By understanding the underlying molecular factor that distinguish the two populations, new molecular target could be discovered for future HER2-based cancer treatment.

The dataset used in this study comes from the BRCA dataset of The Cancer Genome Atlas ("Comprehensive Molecular Portraits of Human Breast Tumours," 2012). TCGA is a publicly available cancer multiomic atlas hosted by the National Institute of Health (Tomczak et al., 2015). Specifically, analysis was focused on clinical outcomes of $HER2^{high/low}$ patient populations, as well as somatic genomic mutations and RNA-Sequencing differential gene expression between $HER2^{high/low}$ patient populations. The analysis shows that $HER2^{high}$ patients are trended to survive better than $HER2^{low}$ patients, presumably due to HER2 being an excellent molecular target for therapies. Interestingly, patients that express high level of HER2 appear to have more 3' mutations in the PIK3CA gene and more 5' mutations in the TP53 gene, although their molecular significance is unknown. This paper offers an observational scaffold for future molecular experiments to operate upon.

Methods

Clinical information, maf object and summarized experiment object was accessed from TCGA under the project name “TCGA-BRCA” with R package TCGAbiolinks (Colaprico et al., 2016). Patient survival time was estimated using days to last followup if the patients are still alive. All patients that do not have valid survival time were filtered out. A bar plot of how many patients fall within each semi-quantitative immunohistochemical HER2 expression level was created using R package ggplot2. A Kaplan–Meier plot was created with patients grouped by the semi-quantitative immunohistochemical level of HER2 expression with R package “survival” (Figure 2).

With maf object, patients that do not have HER2 semi-quantitative immunohistochemical expression level were filtered out using R package maftools (Mayakonda et al., 2018). Patients were then segmented into two groups, low and high HER2 expression, with patients that have 0 or 1+ HER2 expression level in the low group and patients who have 2+ or 3+ HER2 expression level in the high group. A co-oncoplot comparing the low and high HER2 expression groups were created using maftools (Figure 3).

Finally, with summarized experiment object, RNA-Seq counts of normal tissues were filtered out. Patients were then segmented into an old (>58) and young group (<=58) according to the median age when the patients were indexed into the TCGA study. All genes that have less than 10 counts across patient samples were filtered out. Differential gene expression was conducted using DESeq2 with the origin location of cancer and pathologic stage as covariables and age as the main variable (Love et al., 2014). A volcano plot was created using R package “EnhancedVolcano” with adjusted-p value cutoff of 0.05.

Results

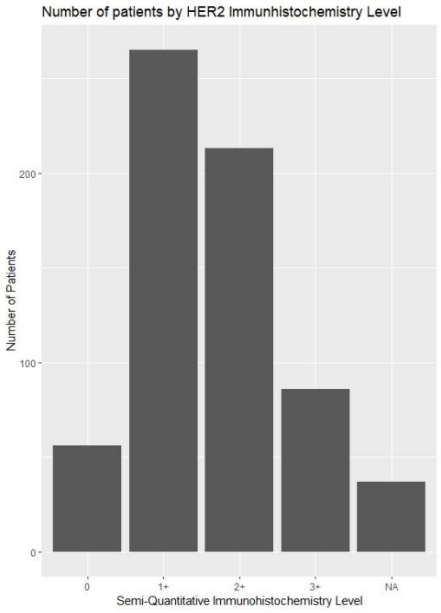

Figure 1. Number of patients in each HER2 semi-quantitative immunohistochemical level. Range from lowest 0 to highest 3+. Unknown is marked with NA.

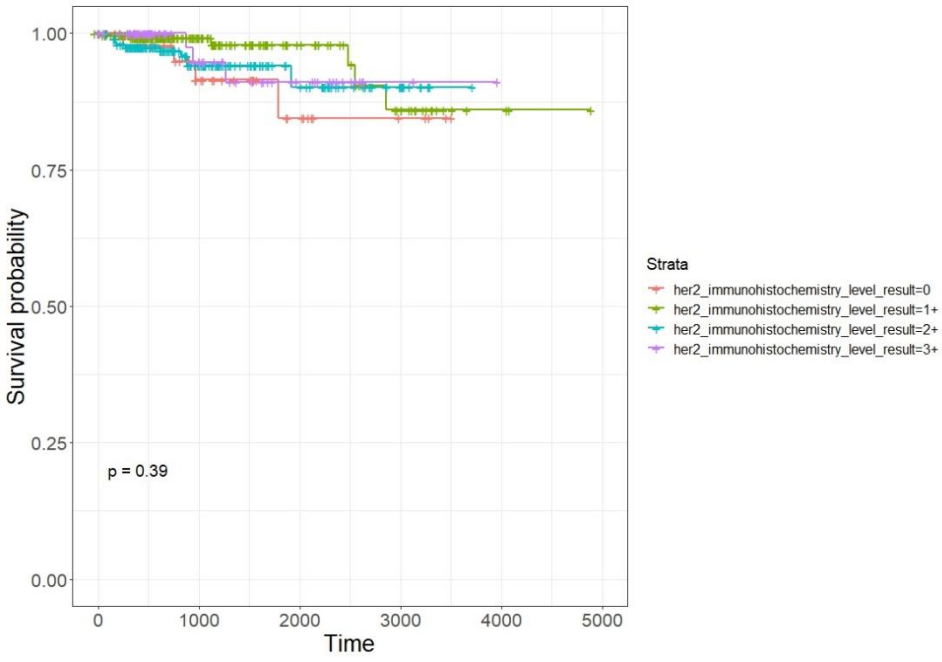

Figure 2. Kaplan–Meier plot of patients based on HER2 immunohistochemical level. The lower-left value indicates the p-value of this study.

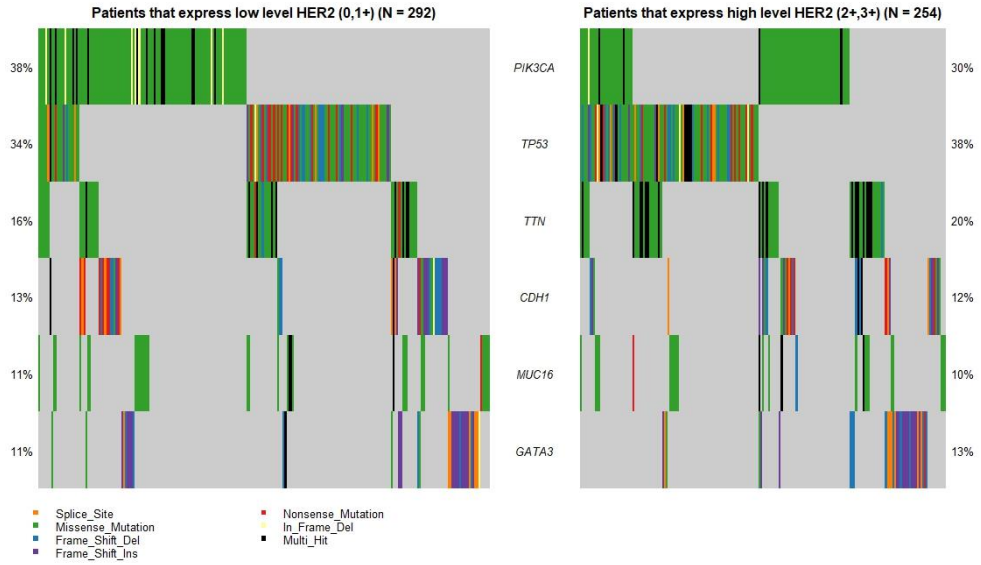

Figure 3. Co-oncoplot of top 6 most-mutated gene among patients with low HER2 expression (immunohistochemical level: 0, 1+; n = 292) and patients with high HER2 expression (immunohistochemical level: 2+, 3+; n = 254). The lower-left illustration indicates what kind of mutation occurs at which location of the gene.

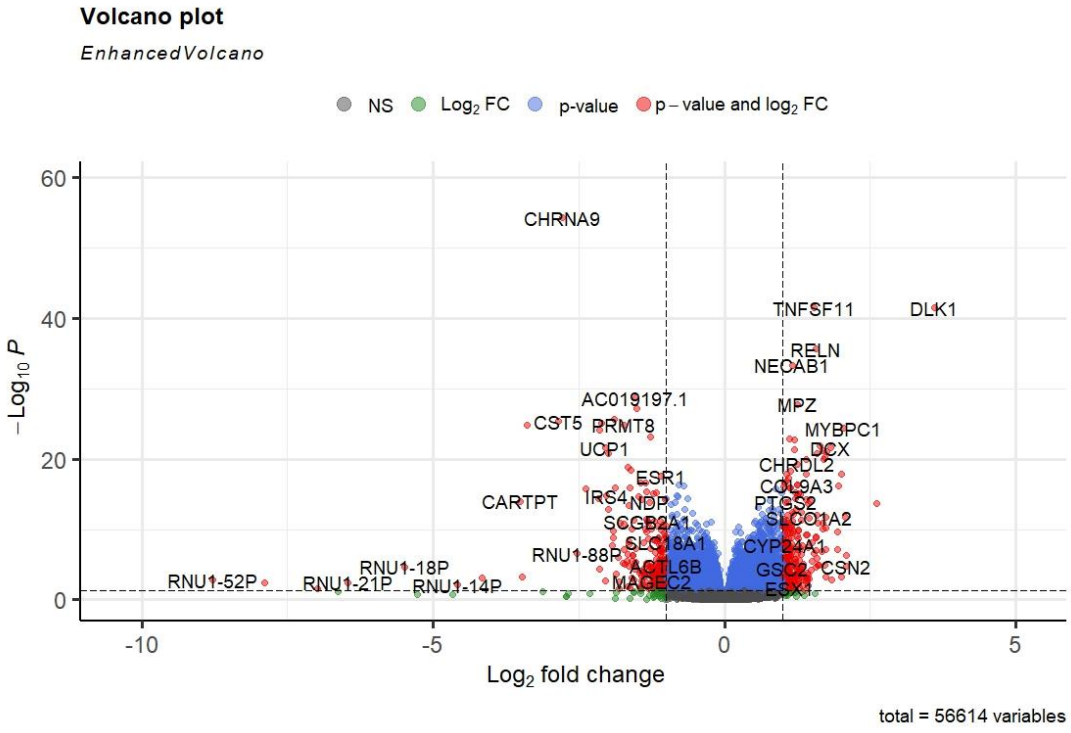

Figure 4. Volcano plot of differentially expressed genes when controlled for age, pathologic stage and cancer origin tissue.

The HER2 semi-quantitative immunohistochemical expression level is approximately normally distributed (Figure 1). Most of the patients are rated 1+ with decreasing tails on either extremes.

The Kaplan-Meier plot does not show a significant p-value with respect to if HER2 immunohistochemical level is statistically significant for cancer survivability (Figure 2). However, it can be clearly observed that there is a trend that higher HER2 expression (2+ or 3+) seems to be correlated with higher survival rate when compared to lower HER2 expression (0 or 1+).

Interestingly, the co-oncoplot between HER2 high-expressing and low-expressing patients show a difference in gene mutation patterns. Patients who express higher level of HER2 appear to have mutations more frequently in the 3' end of PIK3CA when compared to those expressing lower levels of HER2. High HER2-expressing group also seems to experience more mutations in the 5' end of TP53 when compared to the low-expressing HER2 group.

When differential gene expression analysis is performed when controlled for origin of cancer, pathologic stage and age of the patient, there is no transcriptional pattern difference between the young and old group in PIK3CA or TP53 expression.

Discussion

Although not significant in this study, it can be observed that higher expression of HER2 is tentatively correlated with better survival rate of breast cancer, presumably due to HER2 offering a superior molecular target against cancerous cells in line with current literature.

It is interesting that there appears to be a difference in mutational pattern between $HER2^{high/low}$ groups in the gene PIK3CA and TP53. PIK3CA is a gene implicated in cell proliferation and abnormal tissue growth (Canaud et al., 2021). While it is unknown why $HER2^{high}$ tend to have 3' mutations in PIK3CA, it may be plausible that cancerous cells that lack HER2 mutation are more inclined to obtain more deleterious mutations in PIK3CA, since 5' mutations are often more disruptive, in order to be proliferative. TP53, on the other hand, is a critical tumor suppressor (Hu et al., 2021). Again, there is no known molecular evidence regarding the speculation, but it may be speculated that the more deleterious 5' TP53 mutation may be part of the reason why $HER2^{+}$ breast cancer is more aggressive than its $HER2^{-}$ counterpart. Further experimentation is required to understand the significance of the differing mutational pattern in PIK3CA and TP53 among $HER2^{high/low}$ patient populations.

Regardless, this paper offers a unique multiomic view on the underlying molecular differences between $HER2^{high/low}$ breast cancer with the aim to encourage conversation around $HER2^{+/-}$ breast cancer in hope of find better onco-therapies in face of HER2 resistance.

References

Canaud, G., Hammill, A. M., Adams, D., Vikkula, M., & Keppler-Noreuil, K. M. (2021). A review of mechanisms of disease across PIK3CA-related disorders with vascular manifestations. *Orphanet Journal of Rare Diseases*, *16*(1), 306. https://doi.org/10.1186/s13023-021-01929-8

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T. S., Malta, T. M., Pagnotta, S. M., Castiglioni, I., Ceccarelli, M., Bontempi, G., & Noushmehr, H. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Research*, *44*(8), e71–e71. https://doi.org/10.1093/nar/gkv1507

Comprehensive molecular portraits of human breast tumours. (2012). *Nature*, *490*(7418), 61–70. https://doi.org/10.1038/nature11412

Fitzmaurice, C., Dicker, D., Pain, A., Hamavid, H., Moradi-Lakeh, M., MacIntyre, M. F., Allen, C., Hansen, G., Woodbrook, R., Wolfe, C., Hamadeh, R. R., Moore, A., Werdecker, A., Gessner, B. D., Te Ao, B., McMahon, B., Karimkhani, C., Yu, C., Cooke, G. S., … Naghavi, M. (2015). The Global Burden of Cancer 2013. *JAMA Oncology*, *1*(4), 505. https://doi.org/10.1001/jamaoncol.2015.0735

Hu, J., Cao, J., Topatana, W., Juengpanich, S., Li, S., Zhang, B., Shen, J., Cai, L., Cai, X., & Chen, M. (2021). Targeting mutant p53 for cancer therapy: direct and indirect strategies. *Journal of Hematology & Oncology*, *14*(1), 157. https://doi.org/10.1186/s13045-021-01169-0

Krishnamurti, U., & Silverman, J. F. (2014). HER2 in Breast Cancer. *Advances in Anatomic Pathology*, *21*(2), 100–107. https://doi.org/10.1097/PAP.0000000000000015

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

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

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., & Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Research*, *28*(11), 1747–1756. https://doi.org/10.1101/gr.239244.118

Swain, S. M., Shastry, M., & Hamilton, E. (2023). Targeting HER2-positive breast cancer: advances and future directions. *Nature Reviews Drug Discovery*, *22*(2), 101–126. https://doi.org/10.1038/s41573-022-00579-0

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