Comparison of Survival, Gene Expression and Mutation Patterns in Infiltrating Lobular Carcinoma and Infiltrating Ductal Carcinoma

Introduction

The Cancer Genome Atlas, also known as TCGA, is a publicly accessible genomic, transcriptomic, epigenomic and proteomic cancer data source that researchers use to conduct statistical analyses and gain a better understanding of the genetic basis of cancer. Created in a joint effort between the National Cancer Institute and the National Human Genome Research Institute in 2006, the program has generated thousands of samples from over 30 cancer types, leading to tremendous growth in the computational biology and cancer research field. Breast cancer makes one third of all new female cancers per year, making it the most commonly diagnosed form of cancer in women. Within the disease, there are many histological types which are dependent on the location of the cancerous cells, with the most common being invasive cancers including Infiltrating Ductal Carcinoma (IDC) and Infiltrating Lobular Carcinoma (ILC), which make up 80% and 10% of all breast cancers, respectively (Alkabban, F., & Ferguson, T.). Some common risk factors of breast cancers are old age, family genetics, smoking and alcohol use, obesity and radiation exposure. This study aimed to identify any correlations between the histological types of breast cancer patients and survival patterns as well as gene mutation and expression. This was done by conducting statistical analyses to identify differences in gene expression and mutation between IDC and ILC, the most common types of breast cancer, which could potentially help researchers better understand how to develop new therapies specific to histological type to increase breast cancer survival rates.

Methods

Breast cancer clinical and MAF data were accessed from TCGA with the accession code "TCGA-BRCA". To create boxplots, the standard R boxplot function was used along with data from the clinical dataframe regarding the age at diagnosis and histological types of patients. To create Kaplan Meier plots, the R packages survival and survminer were used to perform survival analysis along with vital status data from the clinical dataframe. To create co-oncoplots and other somaticMutations plots such as co-lollipop plots, the R package maftools was used along with the clinical data dataframe. To create volcano plots, the R package EnhancedVolcano was used and differential expression analysis was performed using the DESeq2 R package to generate the DESeq data set. Boolean indexing was used to filter out insignificant data and to identify up regulated and down regulated genes.

Results

It was found that the average age of initial pathological diagnosis for Infiltrating Ductal Carcinoma was under the age of 60 years, while the average age of initial pathological diagnosis for Infiltrating Lobular Carcinoma was approximately 60 years (Figure 1). Additionally, survival analysis showed that patients with Infiltrating Lobular Carcinoma had an overall higher chance of survival over time in comparison to patients with Infiltrating Ductal Carcinoma (Figure 2). The most commonly mutated genes in ILC and IDC were TP53, PIK3CA, TTN, CDH1, GATA3, and MUC16. The biggest difference in mutation percentages between IDC and ILC patients was found in the MUC16 gene, as mutations in MUC16 were recorded in 14% of ILC patients but only 9% of IDC patients (Figure 3). Most mutations found on the TP53 gene in both IDC and ILC were found on the P53 domain (Figure 4). HORMAD1 is the one of the most significantly

upregulated genes in IDC compared to ILC (Figure 5). CSN1S1 is one of the most significantly downregulated genes in IDC compared to ILC (Figure 5).

Figure 1. Boxplot showing that the average age of diagnosis for ILC is somewhat higher than the average age of diagnosis for IDC.

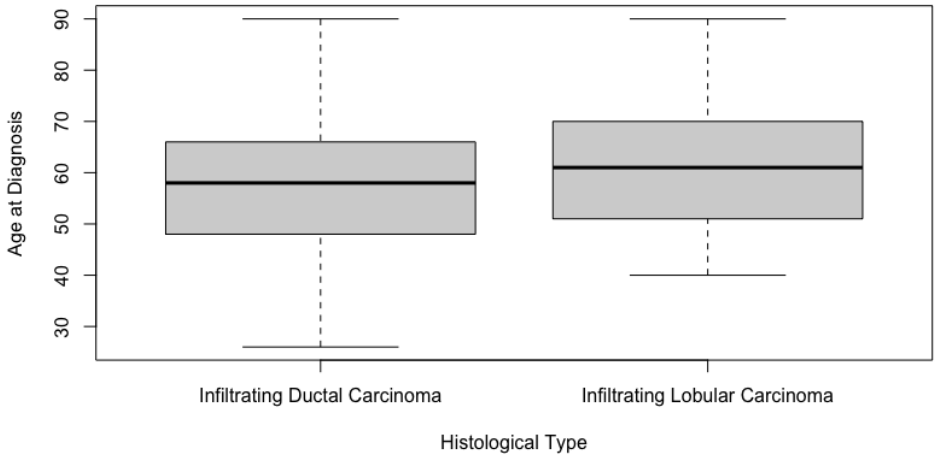


Figure 2. Kaplan-Meier survival plot showing that ILC patients experience significantly decreased survival times in comparison to IDC patients. P-value of 0.019 denotes the statistical significance of the found results.

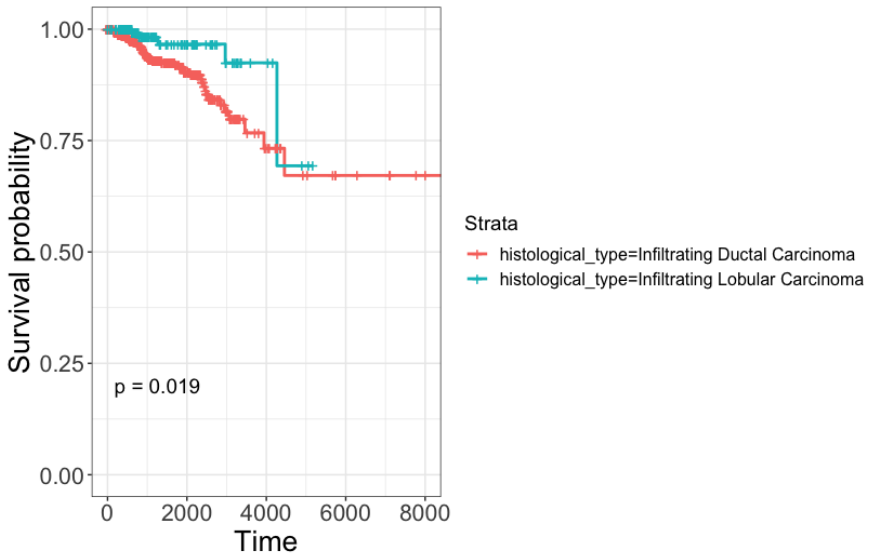


Figure 3. Co-oncoplot showing that the most commonly mutated genes in IDC and ILC had similar percentages of mutation in IDC and ILC patients. The most common type of mutation in both IDC and ILC was missense mutations.

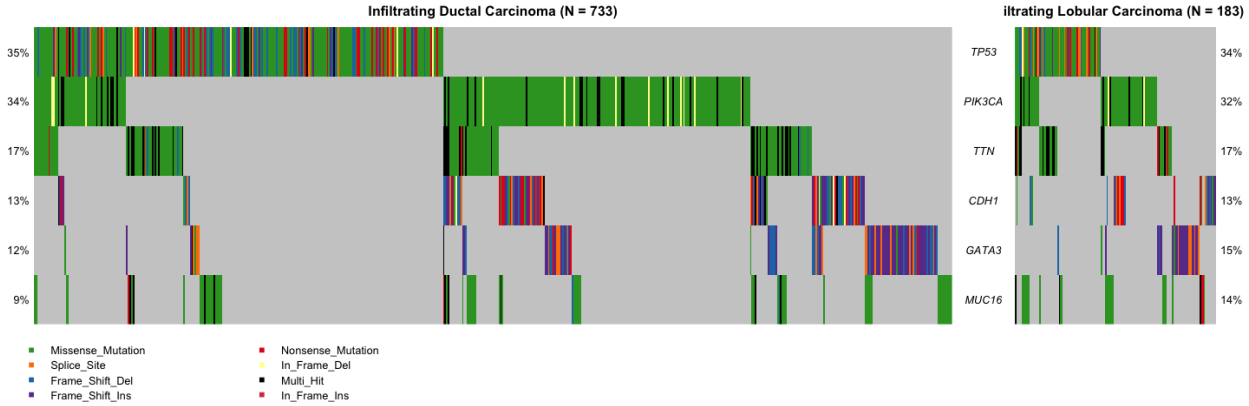


Figure 4. Co-lollipop plot showing that the P53 domain had the highest mutation counts on the TP53 gene for both IDC and ILC. 35.06% of IDC tumors and 34.43% of ILC tumors had mutations on the TP53 gene.

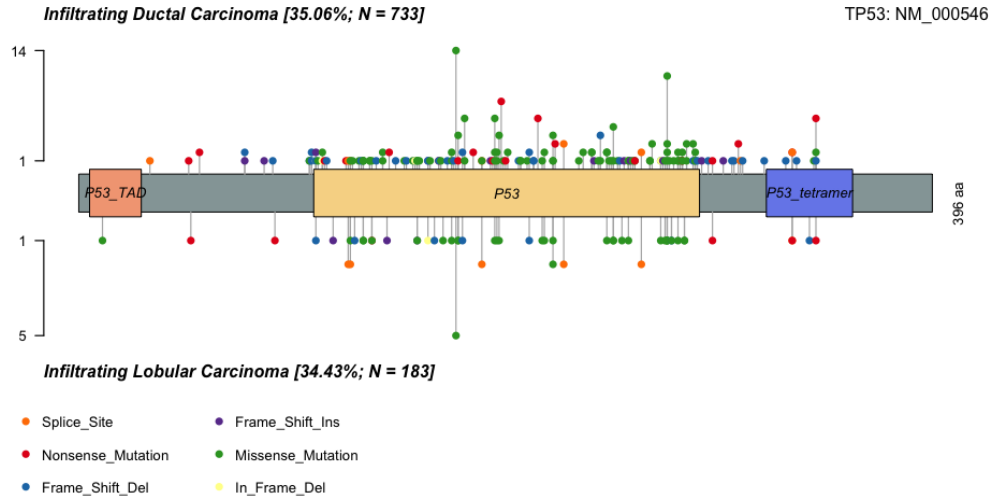
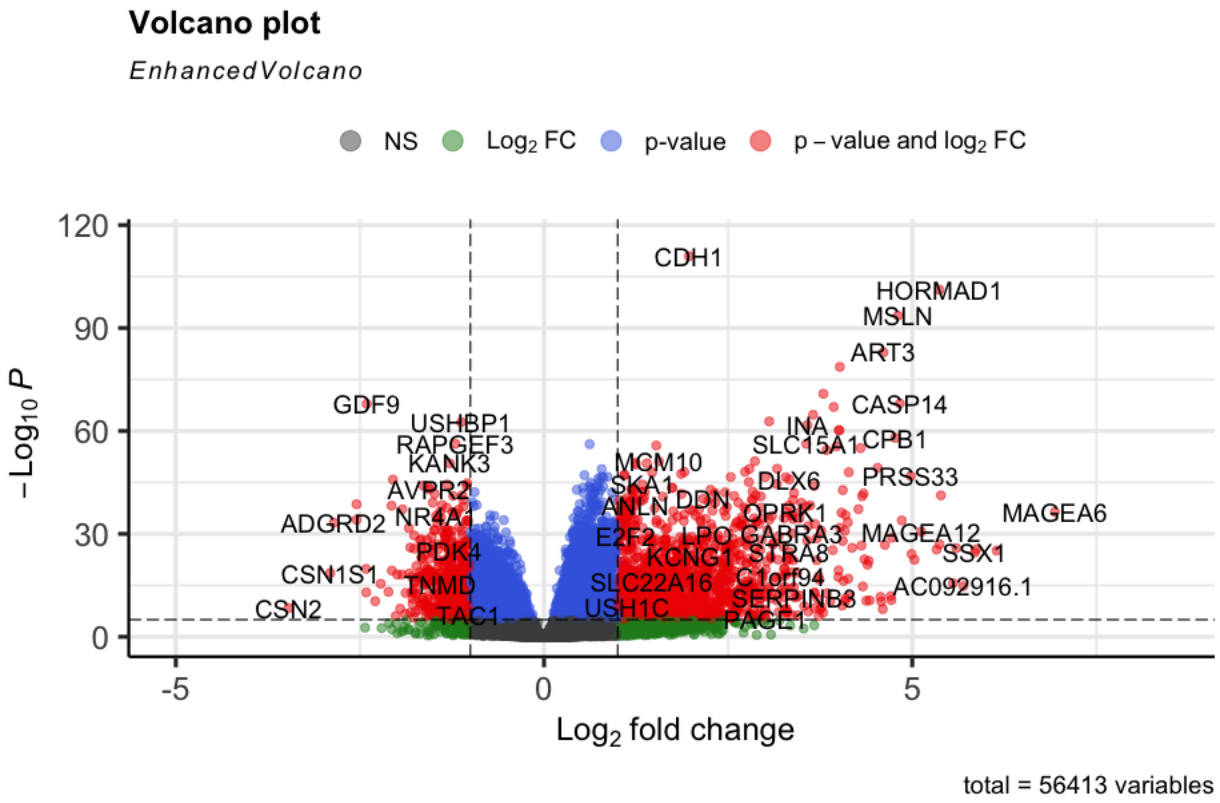


Figure 5. Volcano plot showing that there are more significantly upregulated genes in IDC in comparison to ILC. HORMAD1 is the one of the most significantly upregulated genes in IDC compared to ILC. CSN1S1 is one of the most significantly downregulated genes in IDC compared to ILC.



Discussion

The findings of this study imply that while the average age of diagnosis of ILC patients is above 60 which is slightly higher than that of IDC patients (Figure 1), ILC patients experience significantly decreased survival times in comparison to IDC patients (Figure 2). Similar results were observed in a more long term large population-based study by Chamalidou, C. et al. who found that there was a favorable relative survival rate for patients with ILC compared to IDC during the first 5 years of follow-ups, but survival rate decreased for ILC patients compared to IDC patients after 10 to 15 years of follow-ups (Chamalidou, C. et al., 2021). However, the study

found no significant differences between IDC and ILC in the percent of mutations in TP53, PIK3CA, TTN, CDH1, GATA3, and MUC16 (Figure 3). In Ciriello, G. et al., mutations in TP53 were less frequent in ILC in comparison to IDC, with 8% and 44%, respectively. This difference in results may be due to one of the limitations of this study, which is the poor ratio of IDC to ILC patients in the data (Ciriello, G. el al., 2015). Furthermore, the differential expression analysis suggested that the gene HORMAD1 was upregulated and CSN1S1 was downregulated in IDC in comparison to ILC (Figure 5). Watkins, J. et al. found that the overexpression of HORMAD1 contributes to homologous recombination deficiency and chromosomal instability in triple negative breast cancers, which are estrogen and progesterone receptor negative, and ultimately drives sensitivity of cancer cells to HR targeting therapies (Watkins, J. et al., 2015). Understanding the role of HORMAD1 overexpression in breast cancer patients could allow researchers to develop novel therapies that remain effective on cancer cells. Additionally, Akter, M. et al. found a positive correlation between underregulation of CSN1SI and breast cancer patient survival rates, which contradicts the findings of this study (Akter, M. et al., 2020). For future research, it would be interesting to add in the estrogen receptor and progesterone receptor status as a factor in comparing survival rates of IDC and ILC patients to better understand their role. Although this study was limited to data from under 1000 patients and was unable to find any significant correlation between genomic and transcriptomic characteristics and histological type, future studies with a larger sample size could provide more significant and conclusive findings.

References

Alkabban, F., & Ferguson, T. *Breast cancer*. NCBI Bookshelf. Retrieved March 11, 2023, from https://www.ncbi.nlm.nih.gov/books/NBK482286/

Chamalidou, C., Fohlin, H., Albertsson, P., Arnesson, L.-G., Einbeigi, Z., Holmberg, E., Nordenskjöld, A., Nordenskjöld, B., Karlsson, P., & Linderholm, B. (2021). Survival patterns of invasive lobular and invasive ductal breast cancer in a large population-based cohort with two decades of follow up. *The Breast*, *59*, 294–300. https://doi.org/10.1016/j.breast.2021.07.011

Ciriello, G., Gatza, M., & Beck, A. (2015). Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*.

Mou, M. A., Keya, N. A., Islam, M., Hossain, M. J., Al Habib, M. S., Alam, R., Rana, S., Samad, A., & Ahammad, F. (2020). Validation of CSN1S1 transcriptional expression, promoter methylation, and prognostic power in breast cancer using independent datasets. *Biochemistry and Biophysics Reports*, *24*, 100867. https://doi.org/10.1016/j.bbrep.2020.100867

Watkins, J., Weekes, D., Shah, V., Gazinska, P., Joshi, S., Sidhu, B., Gillett, C., Pinder, S., Vanoli, F., Jasin, M., Mayrhofer, M., Isaksson, A., Cheang, M. C. U., Mirza, H., Frankum, J., Lord, C. J., Ashworth, A., Vinayak, S., Ford, J. M., … Tutt, A. N. J. (2015). Genomic complexity profiling reveals that hormad1 overexpression contributes to

homologous recombination deficiency in triple-negative breast cancers. *Cancer Discovery*, *5*(5), 488–505. https://doi.org/10.1158/2159-8290.cd-14-1092