Identification of biomarkers for breast invasive carcinoma using TCGA data.

Adam Elí Davíðsson (aed5), Margrét Dís Stefánsdóttir (mds11)

Project summary

The project aims to identify potential biomarkers for breast invasive carcinoma using gene expression data obtained from The Cancer Genome Atlas (TCGA). The project will involve clustering, differential expression analysis, and classification to identify the most relevant genes for tissue identification and prognosis. RNA sequencing data will be used to compare gene expression in healthy and cancerous breast tissue and identify highly expressed genes that may serve as potential biomarkers for breast invasive carcinoma prognosis. The project will also explore the relationship between gene expression and increased mortality and open up new opportunities for further research on genes in breast invasive carcinoma tissue and tumours with different grades that may impact prognosis and serve as indicators for early aggressive treatment. This effort to identify potential biomarkers for breast invasive carcinoma using TCGA gene expression data has significant implications for cancer diagnosis, prognosis, and treatment.

Background & preliminary results

The problem this study aims to address is the identification of potential biomarkers for breast invasive carcinoma, a common type of breast cancer that affects women worldwide. Breast invasive carcinoma is associated with high morbidity and mortality rates, and early detection is crucial for effective treatment and improved patient outcomes. Previous studies have identified several potential biomarkers for breast invasive carcinoma, including HER2, estrogen receptor (ER), and progesterone receptor (PR) status. However, these markers have limitations in terms of sensitivity and specificity, and there is a need for more accurate and reliable biomarkers for early diagnosis and prognosis.

To address this gap in knowledge, this study uses gene expression data obtained from The Cancer Genome Atlas (TCGA) to identify new potential biomarkers for breast invasive carcinoma. The study employs clustering, differential expression analysis, and classification to identify the most relevant genes for tissue identification and prognosis. Through RNA sequencing data, the study clusters differences in gene expression between healthy and cancerous breast tissue and identifies highly expressed genes that may be potential biomarkers for breast invasive carcinoma prognosis.

Several studies have already demonstrated the potential of gene expression profiling for identifying biomarkers for cancer diagnosis and prognosis. For instance, a study by Liu et al. (2018) found that a gene expression signature derived from breast cancer tissues could predict patient outcomes and guide treatment decisions. Similarly, a study by Venet et al. (2011) identified a gene expression signature that predicted the risk of distant metastasis in breast cancer patients.

Overall, the identification of new potential biomarkers for breast invasive carcinoma using TCGA gene expression data has significant implications for cancer diagnosis, prognosis, and treatment, and could ultimately lead to improved patient outcomes.

Specific Aims and Research Strategy

Aim 1. Identify potential biomarkers for breast invasive carcinoma prognosis using TCGA gene expression data.

Significance: Breast invasive carcinoma is the most common type of breast cancer and identifying potential biomarkers for prognosis can lead to better treatment planning and ultimately improve patient outcomes.

Innovation: This study will utilise RNA sequencing data to identify potential biomarkers for breast invasive carcinoma prognosis, which is an innovative approach that can provide more accurate predictions.

Approach: The TCGAbiolinks package will be used to retrieve and preprocess data from the TCGA database. Principal component analysis will be performed to assess differences in gene expression patterns between normal and cancerous tissue. Differential expression analysis and an Elastic Net classifier model will be used to identify the most relevant genes for predicting whether tissue samples were from diseased or normal healthy tissue. The selected genes are then hierarchically clustered based on their predictive qualities and highlighted in a heatmap.

Aim 2. Determine the correlation between TCGA gene biomarkers and survival rates in breast invasive carcinoma patients.

Significance: Understanding the correlation between gene expression and increased mortality can help identify potential biomarkers for early aggressive treatment.

Innovation: This study will focus on the correlation between gene biomarkers and survival rates in breast invasive carcinoma patients, which is an innovative approach that can improve treatment planning and patient outcomes.

Approach: Analyse the correlation between TCGA gene biomarkers and survival rates in breast invasive carcinoma patients using Kaplan-Meier plots.

Aim 3. Investigate the potential of identified biomarkers for breast invasive carcinoma tumours with different grades that may affect prognosis.

Significance: This study would provide insights into the potential use of identified biomarkers for tumours with different grades, which can lead to better treatment planning and improved patient outcomes.

Innovation: This study intends to investigate the potential of identified biomarkers for breast invasive carcinoma tumours with different grades, which is an innovative approach that can improve treatment planning and patient outcomes.

Approach: The identified potential biomarkers will be further analysed in breast invasive carcinoma tumours with different grades to investigate their potential use as indicators for early aggressive treatment.

References

1. American Cancer Society. (2022). Breast Cancer. https://www.cancer.org/cancer/breast-cancer.html

2. National Cancer Institute. (2021). Breast Cancer Treatment (PDQ®)–Patient Version. https://www.cancer.gov/types/breast/patient/breast-treatment-pdq

3. European Medicines Agency. (2016). Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-use-pharmacogenetic-methodologies-pharmacokinetic-evaluation-medicinal-products_en.pdf

4. Zhang, J., & Shu, C. (2017). Biomarkers of breast cancer. Handbook of Experimental Pharmacology, 249, 93-108. https://doi.org/10.1007/164_2017_25

5. Shyr, Y., & Kim, K. (2014). Overview of RNA sequencing and applications. In RNA sequencing (pp. 1-23). Springer. https://doi.org/10.1007/978-1-4939-1392-4_1

6. Wang, S. Y., & Li, H. (2017). Correlation of EGFR gene mutations with prognosis in patients with breast invasive carcinoma. Oncology Letters, 13(3), 1573-1576. https://doi.org/10.3892/ol.2017.5622

7. World Health Organization. Breast Cancer: Prevention and Control. https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

8. The Cancer Genome Atlas: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

9. Liu, S., Shen, D., Konecny, G. et al. Prognostic biomarker discovery in breast cancer by gene expression profiling. Breast Cancer Res Treat 173, 247–257 (2018). https://doi.org/10.1007/s10549-018-4838-1

10. Venet, D., Dumont, J.E., Detours, V. Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome. PLoS Comput Biol 7, e1002240 (2011). https://doi.org/10.1371/journal.pcbi.1002240