Gene Expression Analysis and Interpretation of Patients with ERBB2 amplified Breast Cancer

Tilman Schaefer 23206466

Introduction

According to the International Agency for Research on Cancer, breast cancer is the most common cancer overall, accounting for about 12% of all cancer cases [7]. In approx. 25-30% of invasive breast cancers the ERBB2 oncogene is overexpressed [1], which has also been linked to the promotion of breast cancer invasion and metastasis, resulting in poor patient survival rates [2,3].

ERBB2 belongs to the ERBB family of genes that encode a member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases (RTKs) [3]. Although this protein has no ligand binding domain of its own and therefore cannot bind growth factors, it does bind with other EGF receptors and activates downstream signalling pathways such as MAPK and PI3K/Akt to the effect of promoting cell proliferation and suppressing apoptosis.

While the overexpression of ERBB2 has been established as a reliable biomarker for the diagnosis, treatment and prognosis of breast cancer, many of the underlying processes such as tumour progression and resistance to treatment are still not well understood. It is therefore important to understand the factors that contribute to therapy resistance of ERBB2-positive breast cancer tumours and to identify other genetic or transcriptomic factors in order to identify novel therapeutic strategies to overcome resistance.

With the advent of RNA sequencing techniques (RNA-Seq) using next-generation sequencing it is now possible to detect and sample the presence and quantity of RNA in a living organism, which captures the organisms dynamic pool of RNAs or transcriptome at a point-in-time [6]. The data obtained from RNA-Seq can then be used to perform statistical analyses such as differential gene expression to assess a gene's expression levels, or pathway enrichment analysis to identify deregulated signalling pathways.

In this study I am going to investigate to what extend overexpression of genes occurs in patients with ERBB2+ breast cancer, and whether functional pathways are overrepresented in these patients. This could provide insights into other contributing factors that promote/supress

tumour progression or enhance/inhibit therapy resistance, ultimately leading to a better understanding and the development of alternative therapy targets.

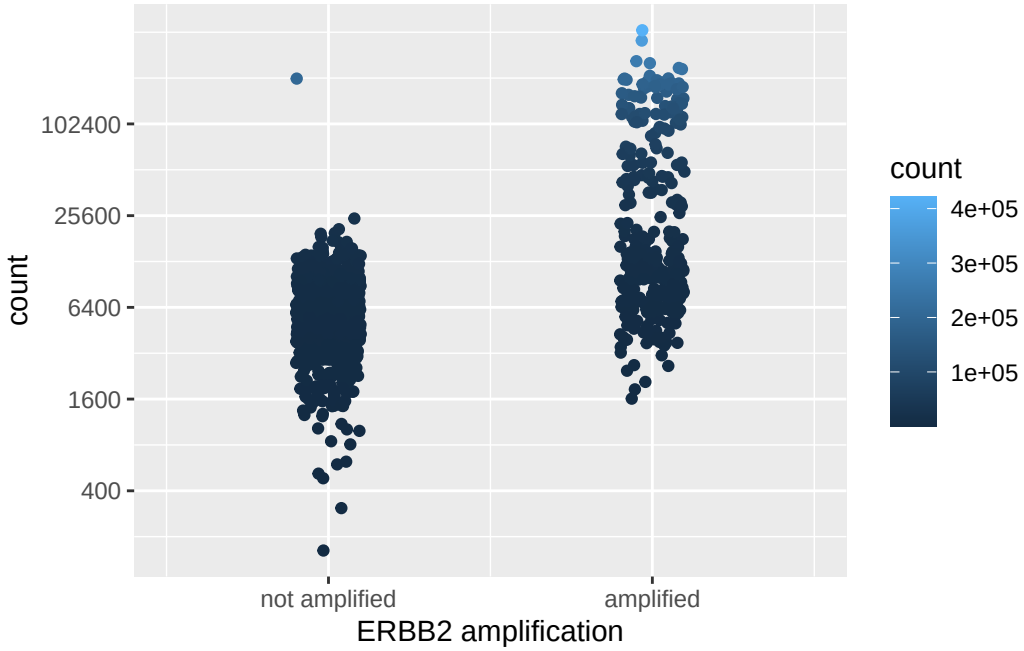
Methods

For this investigation I performed a Differential Gene Expression analysis of patients with BRCA, using the DESeq2 R-package from BiocManager. The patient data was obtained from cBioPortal and contains a dataset with 1082 observations [9]. For the analysis I divided the patient's data into two groups, one group with an amplified level of ERBB2 (CNA level > 0) and the other one with an expression level <= 0. 328 patients exhibited an amplified ERBB2 level, whereas 754 patients did not. A KEGG Enrichment Analysis was subsequently performed using the clusterProfiler library from BiocManager. Lastly, I performed a Principal Component Analysis.

Results

Data Preparation

For data preparation I filtered the RNASeq dataset to only contain read counts of >= 10.



In order to assess the overall quality of the filtered data I evaluated the normalised counts for the two patient groups. For this I plotted the distribution of counts of the ERBB2 gene. For non-amplified ERBB2 patients we can observe a generally lower count compared to ERBB2 amplified patients, together with a smaller distribution.

Top 10 Differentially Expressed Genes Ranked by Fold Change

The aim of differential expression analysis is to discover significant changes in expression levels of genes. The difference in expression levels from a control group can then function as an indicator of a common underlying factor in the development of a disease. It also serves as a pre-requisite for performing a Pathway Enrichment Analysis. In this section I obtained the top 10 genes that were differentially expressed based on the observed log2 fold change. Since the log2 fold change can be positive (increased expression) or negative (decreased expression), I retrieved the genes with the top absolute value of the log2 fold change and ordered them accordingly. The result is summarised in Table 1.

Table 1: Top 10 differentially expressed genes

Gene Name	baseMean	log2FoldChange	pvalue	padj
CSN2	18.331467	-4.540222	0.0000000	0.0000003
SPANXA2	2.312353	4.359869	0.0000000	0.0000000
GAGE12D	10.123610	4.322133	0.0000001	0.0000004
SPANXC	2.410721	4.138504	0.0000000	0.0000000
GAGE2B	1.503136	4.067200	0.0041139	0.0099585
SMR3B	60.736503	-3.768110	0.0000000	0.0000000
CSN3	47.977244	-3.728425	0.0000000	0.0000000
LALBA	26.227637	-3.495534	0.0000012	0.0000074
FAM9C	1.673085	3.393388	0.0000000	0.0000000
GAGE4	4.394711	3.332880	0.0000007	0.0000046

Pathway Enrichment Analysis

The DGE analysis performed before allows to subsequently identify more specific groups or categories of genes that together play a role in signalling pathways and intra-cellular processes. Figure 1 shows the result of the pathway enrichment analysis in form of a dotplot.

Figure 1 combines a number of factors from the result of the PEA in order to visualise the relevance of each pathway:

1. The adjusted p-value (p.adjust). Smaller p-values are coloured red whereas higher p-values are coloured blue

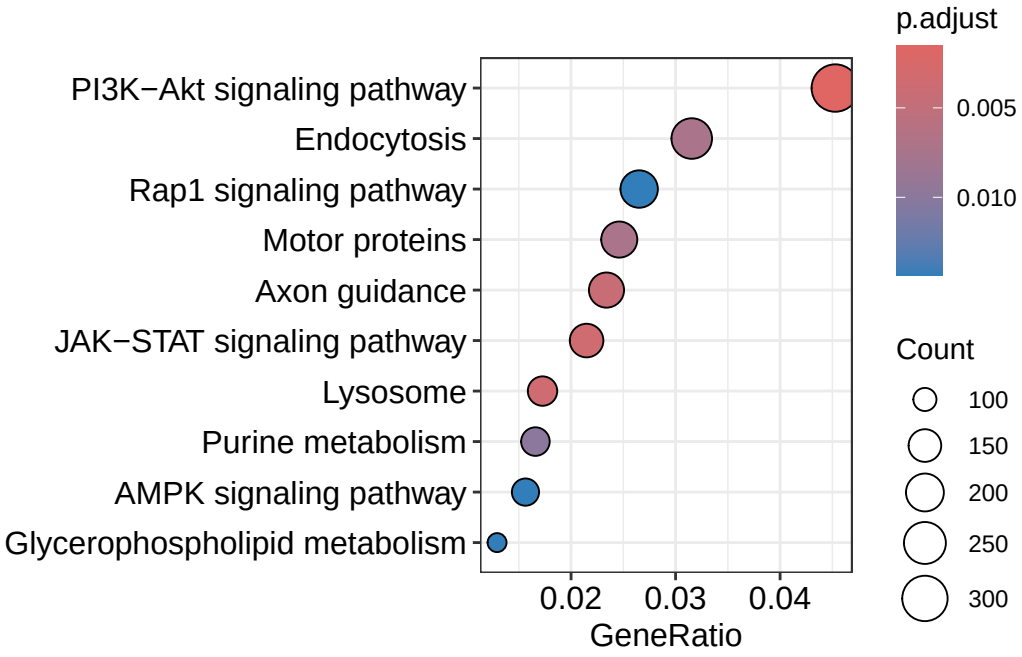


Figure 1: Pathway Enrichment Analysis

2. GeneRatio: This is the ratio of enrichment genes over the of count of pathway genes
3. Count: the number of genes that belong to a given gene-set

The two pathways that scored the lowest p-values are P13K-Akt (p-value: 0.0015329) and JAK-STAT (0.0036038). The P13K pathway is one of the most important intracellular pathways, which regulates cell growth, survival, metabolism, and angiogenesis. It is also believed to be deregulated in a wide spectrum of human cancers [4]. The JAK-STAT pathway on the other hand plays an important role in the regulation of the immune system [5].

PCA

The purpose of conducting a PCA is to reduce the number of dimensions in the data, which facilitates the recognition of patterns and the visualisation of the data. For this I first transformed the data using the Variance Stabilizing Transformation (VST) and then obtained the PCA via the `prcomp` function.

The PCA plot is shown in Figure 2. In our case the PCA has not resulted in a clear clustering of data.

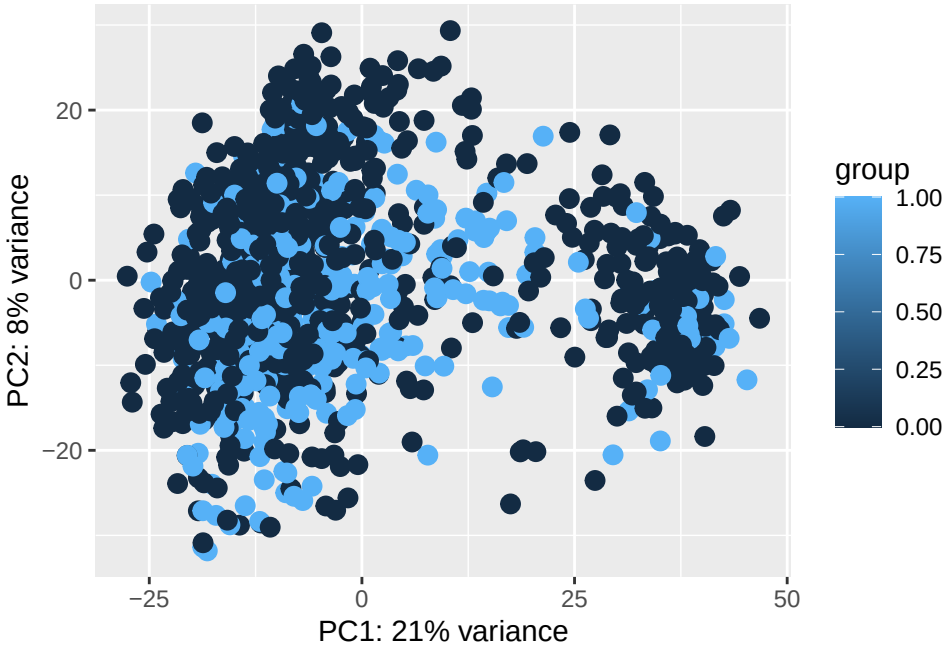


Figure 2: PCA analysis

Discussion

This study investigated in patients with BRCA that exhibited an amplified expression of the ERBB2 oncogene. For this I performed a Differential Gene Expression analysis followed by a Pathway Enrichment analysis.

The result obtained from the DGE analysis revealed a number of genes with elevated expression levels that belong to a group of genes called cancer-testis antigens (CTAs). CTAs are a category of tumor antigens with normal expression restricted to male germ cells in the testis but not in adult somatic tissues [13]. Among these genes are SPANX-A/C and GAGE, which have shown to be induced in breast cancer cells and promote cancer cell invasion [8].

The PEA exhibited some interesting results, indicating a significant deregulation of 2 major pathways, P13K-Akt and JAK-STAT, which both play a role in cell growth, survival, and the immune system. For the remainder of this discussion I will focus on the JAK-STAT pathway since recent advances in the study of this pathway have led to a better understanding of its role in the progression and chemoresistance of BRCA [10].

The JAK-STAT pathway has been identified to regulate embryonic development and is involved in the control of processes such as stem cell maintenance, blood cell formation (haematopoiesis) and the inflammatory response [11]. Suppression of the JAK-STAT pathway can therefore be a promising target for the treatment of BRCA. A number of JAK inhibitors exists to date,

including ruxolitinib, a drug approved by the United States Food and Drug Administration used for the treatment of myelofibrosis [12]. Although a pre-clinical trial with ruxolitinib resulted in some suppression of cell proliferation and tumour growth in tamoxifen-resistant breast cancer cells, overall the observed level of inhibition remained minimal [12]. Further development of JAK inhibitors is therefore an area that will require more research.

References

[1] Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., & Ullrich, A. (1989). Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science (New York, N.Y.), 244(4905), 707–712. https://doi.org/10.1126/science.2470152

[2] Tan, M., Yao, J., & Yu, D. (1997). Overexpression of the c-erbB-2 gene enhanced intrinsic metastasis potential in human breast cancer cells without increasing their transformation abilities. Cancer research, 57(6), 1199–1205.

[3] Uckun, F. M., & Qazi, S. (2023). Upregulated Expression of ERBB2/HER2 in Multiple Myeloma as a Predictor of Poor Survival Outcomes. International journal of molecular sciences, 24(12), 9943. https://doi.org/10.3390/ijms24129943

[4] Yang, J., Nie, J., Ma, X. et al. (2019). Targeting PI3K in cancer: mechanisms and advances in clinical trials. Mol Cancer 18, 26. https://doi.org/10.1186/s12943-019-0954-x

[5] Seif, F., Khoshmirsafa, M., Aazami, H. et al. (2017). The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. Cell Communication and Signaling 15, 23.

[6] Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics, 10(1), 57–63. https://doi.org/10.1038/nrg2484

[7] International Agency for Research on Cancer, Cancer fact sheets (2020), Available at: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf

[8] Maine E. A., Westcott J. M., Prechtl A. M., Dang T. T., Whitehurst A. W., Pearson G. W. The cancer-testis antigens SPANX-A/C/D and CTAG2 promote breast cancer invasion (2016). Oncotarget. 7: 14708-14726. Retrieved from https://www.oncotarget.com/article/7408/text/

[9] cBioPortal for Cancer Genomics. Breast Invasive Carcinoma TCGA PanCancer data. Retrieved from https://www.cbioportal.org/study/summary?id=brca_tcga_pan_can_atlas_2018

[10] Ma, Jh., Qin, L. & Li, X. Role of STAT3 signaling pathway in breast cancer. Cell Commun Signal 18, 33 (2020). https://doi.org/10.1186/s12964-020-0527-z

[11] Thomas, S., Snowden, J., Zeidler, M. et al. The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. Br J Cancer 113, 365–371 (2015). https://doi.org/10.1038/bjc.2015.233

[12] Kim, J. W., Gautam, J., Kim, J. E., Kim, J. A., & Kang, K. W. (2019). Inhibition of tumor growth and angiogenesis of tamoxifen-resistant breast cancer cells by ruxolitinib, a selective JAK2 inhibitor. Oncology letters, 17(4), 3981–3989. https://doi.org/10.3892/ol.2019.10059

[13] Scanlan, M. J., Gure, A. O., Jungbluth, A. A., Old, L. J., & Chen, Y. T. (2002). Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. Immunological reviews, 188, 22–32. https://doi.org/10.1034/j.1600-065x.2002.18803.x