

# Transcriptomic Differences Between HER2-Amplified and Non-Amplified Breast Tumours in TCGA-BRCA

Hanna Dmochowska

Date: 29.11.2025

Word Count: 2971

## Background

Breast cancer is a heterogeneous disease encompassing multiple molecular subtypes with distinct molecular drivers, clinical outcomes, and treatment strategies (Perou et al., 2000). One clinically important subtype is defined by amplification and overexpression of the human epidermal growth factor receptor 2 (*HER2/ERBB2*). HER2-positive tumours constitute approximately 15-20% of breast cancers and are typically associated with high proliferative activity, increased recurrence risk, and poorer prognosis in the absence of targeted therapy (Slamon et al., 1987; Sorlie et al., 2001). The introduction of HER2-targeted agents, such as trastuzumab, has substantially improved outcomes, yet HER2 amplification remains a defining biological event that shapes tumour behaviour (Baselga & Swain, 2009).

Although the oncogenic role of HER2 is well established, less is known about the broader transcriptional programs that distinguish HER2-amplified tumours from other subtypes in large patient cohorts. Understanding these downstream changes is important because HER2 signalling intersects with multiple cellular processes, including proliferation, metabolism, and RNA regulation (Hynes & Lane, 2005). Transcriptome-wide differential expression analysis provides an unbiased approach to identify these processes and to characterise the molecular phenotype associated with *ERBB2* amplification.

In this project, publicly available RNA-sequencing data from TCGA-BRCA were analysed to compare the gene expression profiles of HER2-amplified (HER2+) and non-amplified (HER2-) breast tumours. The aim was to address a central question commonly explored in breast cancer research: which genes and biological pathways are differentially expressed in HER2-amplified tumours, and how do these patterns align with established knowledge of HER2-driven biology? Investigating this question using a large, well-characterised dataset such as TCGA supports a systematic evaluation of the downstream transcriptional consequences of *ERBB2* amplification. This approach allows known mechanisms, such as increased proliferative signalling, to be identified through unbiased analysis, while also highlighting additional pathways that may contribute to tumour behaviour or represent potential therapeutic vulnerabilities. These considerations provide the rationale for the differential expression, pathway enrichment, and exploratory analyses performed in this report.

## Methods

All analyses were performed in R using data from the TCGA Breast Invasive Carcinoma Pan-Cancer Atlas 2018 cohort accessed via cBioPortal (TCGA, 2018). The cohort contains 1,084 primary breast tumour samples with matched genomic, transcriptomic, and clinical information. Three processed datasets were provided for the assignment:

1. **RNA-seq expression** (`data_mrna_seq_v2_rsem.txt`) - A gene × sample matrix containing RSEM-normalised expression values. Rows represent genes (with gene symbols) and columns represent individual tumour samples identified by TCGA barcodes.
2. **Clinical data** (`data_clinical_patient.txt`) - Contains one row per patient, including demographic information, tumour subtype annotations, and overall survival time/status.
3. **Copy-number data** (`data_cna.txt`) - A gene × sample matrix containing gene-level log<sub>2</sub> copy-number changes. Rows correspond to genes, columns correspond to tumour samples using the same TCGA identifiers as the RNA-seq matrix.

TCGA sample barcodes were standardised to the first 12 characters to ensure consistent matching across all datasets. HER2 status was defined using *ERBB2* copy-number values: *ERBB2* > 0 classified as HER2-amplified, and *ERBB2* = 0 as non-amplified. Only samples with matched RNA-seq and CNA data were retained.

Differential expression was performed using DESeq2 (Love, Huber & Anders 2014). PCA was run on VST-transformed counts, and a heatmap of the top 500 variable genes was generated with pheatmap (Kolde 2019). Visualisations were created with ggplot2 (Wickham, 2016). Pathway analysis used clusterProfiler (Xu et al., 2024) and ReactomePA (Yu & He 2016). Survival modelling employed glmnet (Friedman, Hastie and Tibshirani, 2010) and the survival package (Therneau, 2022).

All analysis scripts are provided in the GitHub repository.

## Results

### 1. Differential Expression Between HER2-Amplified and Non-Amplified Tumours

To quantify transcriptomic changes associated with *ERBB2* amplification, a differential expression analysis using DESeq2 was performed. After barcode harmonisation and removal of samples without matched CNA values, 1,068 tumour samples were included, comprising 328 HER2-amplified (Amp) and 740 non-amplified (NotAmp) tumours.

A total of 9,725 genes were significantly differentially expressed at  $\text{padj} < 0.05$ , of which 4,163 were upregulated and 5,562 were downregulated in HER2-amplified tumours. The most statistically significant upregulated genes were *ERBB2*, *PSMD3*, *STARD3*, *C17orf37*, *GRB7*, and *PGAP3*, all located within the 17q12 HER2 amplicon region. This confirms the reliability of the HER2 classification and reflects the expected genomic co-amplification pattern.

The genes with the largest positive fold changes included cancer-testis antigens such as *SPANXA2*, *SPANXC*, *GAGE12D*, and *GAGE2B*, suggesting subtype-specific expression programmes beyond the core HER2 locus. The most strongly downregulated genes included *CSN2* and *CSN3*, which encode caseins typically expressed in differentiated mammary luminal cells, consistent with the more proliferative, less differentiated phenotype of HER2-driven tumours.

Overall, the differential expression profile reflects two major biological patterns: (i) marked upregulation of cell-cycle and mitotic regulators, and (ii) downregulation of genes involved in differentiated epithelial functions and RNA quality-control pathways. These findings align with known HER2 biology and provide the basis for downstream pathway enrichment analyses.

### 2. Top 10 Differentially Expressed Genes Ranked by Fold Change

To identify the most strongly dysregulated genes between HER2-amplified and non-amplified tumours, all significant DEGs were ranked by absolute log<sub>2</sub> fold change. The top

10 genes are shown in Table 1. This list includes both strongly upregulated and strongly downregulated genes in HER2+ tumours.

Several of the highest positive fold-change genes (*SPANXA2*, *SPANXC*, *GAGE12D*, *GAGE2B*, *GAGE4*) belong to the cancer-testis antigen (CTA) family, which is typically expressed in immune-privileged tissues (e.g., testis) and aberrantly reactivated in certain cancers. Their upregulation suggests activation of germline-like transcriptional programs in a subset of HER2+ tumours.

The most downregulated genes (*CSN2*, *CSN3*, *LALBA*) encode proteins normally expressed in differentiated mammary epithelial cells (including caseins and lactalbumin). Their reduced expression is consistent with the less differentiated, highly proliferative phenotype of HER2-driven tumours.

	<b>baseMean</b>	<b>log2Fold Change</b>	<b>IfcSE</b>	<b>stat</b>	<b>pvalue</b>	<b>padj</b>	<b>gene</b>
CSN2	18.527933	-4.563540	0.8454341	-5.397866	0.0000001	0.0000006	CSN2
SPANXA2	2.341483	4.351822	0.6355704	6.847112	0.0000000	0.0000000	SPANXA 2
GAGE12D	10.260942	4.296769	0.7966197	5.393751	0.0000001	0.0000006	GAGE12 D
SPANXC	2.429928	4.172454	0.5238789	7.964539	0.0000000	0.0000000	SPANXC
GAGE2B	1.522606	4.056978	1.4127677	2.871652	0.0040833	0.0104342	GAGE2B
CSN3	47.147610	-3.709707	0.4923039	-7.535400	0.0000000	0.0000000	CSN3
FAM9C	1.694494	3.374503	0.2976388	11.337578	0.0000000	0.0000000	FAM9C
PNMT	165.00059 6	3.309590	0.1793591	18.452313	0.0000000	0.0000000	PNMT
GAGE4	4.453785	3.305694	0.6709006	4.927249	0.0000008	0.0000054	GAGE4
LALBA	22.783840	-3.290685	0.7292141	-4.512646	0.0000064	0.0000339	LALBA

**Table 1. Top 10 differentially expressed genes ranked by absolute log2 fold change.**

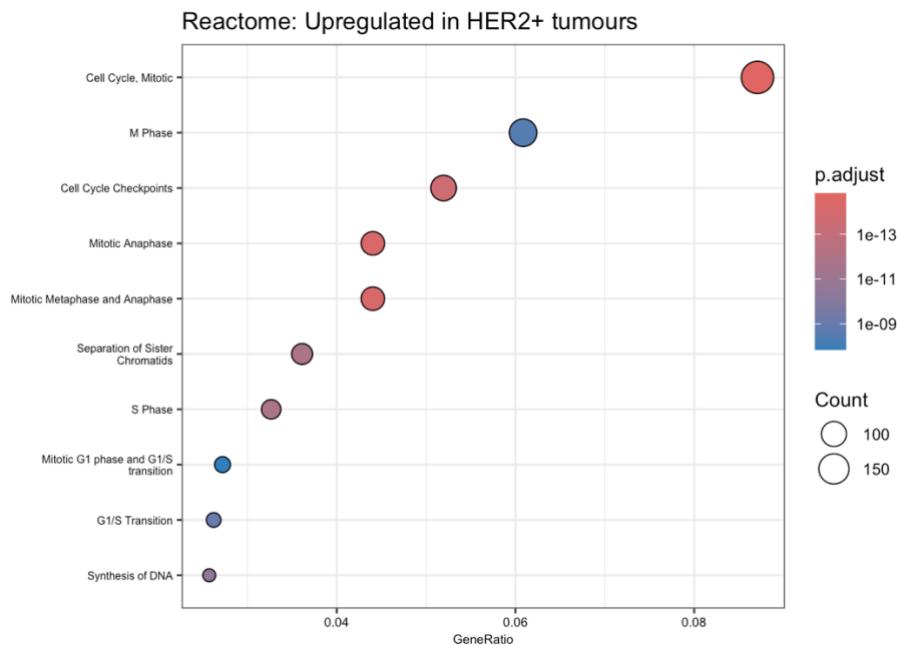
log2FC denotes log2 fold change (positive = upregulated in HER2-amplified tumours; negative = downregulated). padj refers to the Benjamini-Hochberg adjusted p-value. “Direction in HER2+” indicates whether the gene is upregulated or downregulated relative to HER2- tumours.

### 3. Reactome Pathway Enrichment

#### Upregulated pathways in HER2-amplified tumours

Reactome pathway enrichment of genes upregulated in HER2-amplified tumours identified a coherent set of pathways involved in cell-cycle progression and mitotic control (Figure 1). The most significantly enriched terms included *Cell Cycle*, *Mitotic* and *M Phase*, both of which appeared with the highest GeneRatio values and the lowest adjusted p-values. Additional enriched pathways reflected distinct steps of mitosis and DNA replication, including *Cell Cycle Checkpoints*, *Mitotic Anaphase*, *Mitotic Metaphase* and *Anaphase*, *Separation of Sister Chromatids*, *S Phase*, *Mitotic G1 Phase* and *G1/S Transition*, *G1/S Transition*, and *Synthesis of DNA*.

Together, these pathways depict a coordinated upregulation of genes responsible for DNA synthesis, checkpoint signalling, spindle organisation, and chromosome segregation, consistent with increased proliferative activity in HER2-amplified tumours.



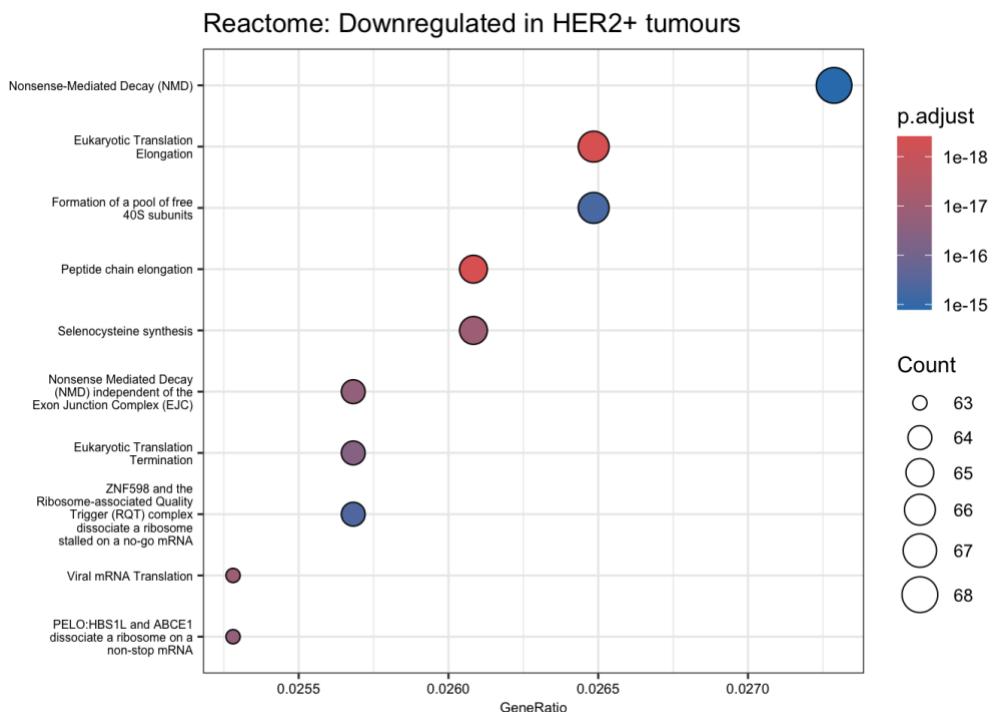
**Figure 1. Reactome enrichment of upregulated genes in HER2-amplified tumours.**

Bubble plot showing the top enriched Reactome pathways among upregulated genes. x-axis: GeneRatio (proportion of upregulated DEGs mapping to each pathway) y-axis: Pathway name. Bubble size: number of genes in the pathway. Bubble colour: adjusted p-value (red = most significant)

## Downregulated pathways in HER2-amplified tumours

Enrichment analysis of genes downregulated in HER2-amplified tumours revealed significant over-representation of pathways involved in mRNA surveillance, translation, and ribosomal function (Figure 2). The strongest enrichment was observed for *Nonsense-Mediated Decay (NMD)* and *Eukaryotic Translation Elongation*, which showed the largest bubbles and the lowest adjusted p-values. Several additional pathways were related to specific translational processes, including *Formation of a Pool of Free 40S Subunits*, *Peptide Chain Elongation*, *Selenocysteine Synthesis*, *NMD independent of the Exon Junction Complex (EJC)* and *Eukaryotic Translation Termination*.

Further enriched terms, such as *ZNF598 and the RQT Complex Dissociate a Ribosome Stalled on a No-Go mRNA*, *Viral mRNA Translation*, and *PELO:HBS1L and ABCE1 Dissociate a Ribosome on a Non-Stop mRNA*, indicate downregulation of pathways responsible for resolving aberrant translation events and stalled ribosomes.



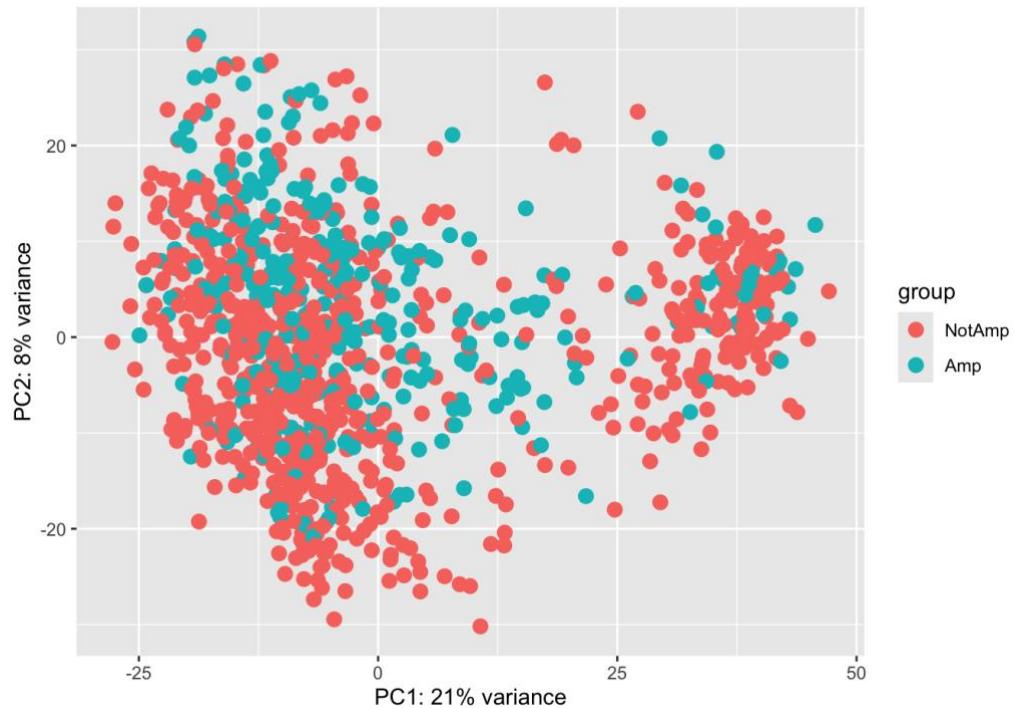
**Figure 2. Reactome enrichment of downregulated genes in HER2-amplified tumours.**

Bubble plot showing enriched Reactome pathways among downregulated genes. x-axis: GeneRatio. y-axis: Pathway name. Bubble size: number of genes contributing to pathway. Bubble colour: adjusted p-value

## 4. Principal Component Analysis (PCA)

Principal component analysis was performed on variance-stabilised (VST) expression values to assess global transcriptional variation across HER2-amplified and non-amplified tumours. The first two principal components accounted for 21% (PC1) and 8% (PC2) of the variance in the dataset (Figure 3).

Although HER2-amplified and non-amplified samples are distributed across the full PCA space, a subtle separation is visible along PC1, with HER2-amplified tumours showing a slight shift relative to non-amplified cases. However, the two groups show extensive overlap, indicating that *ERBB2* amplification contributes to, but does not solely define, the dominant axes of transcriptomic variation in the TCGA-BRCA cohort. This pattern is consistent with the molecular heterogeneity of breast cancer, in which HER2 amplification co-occurs with diverse transcriptional subtypes.



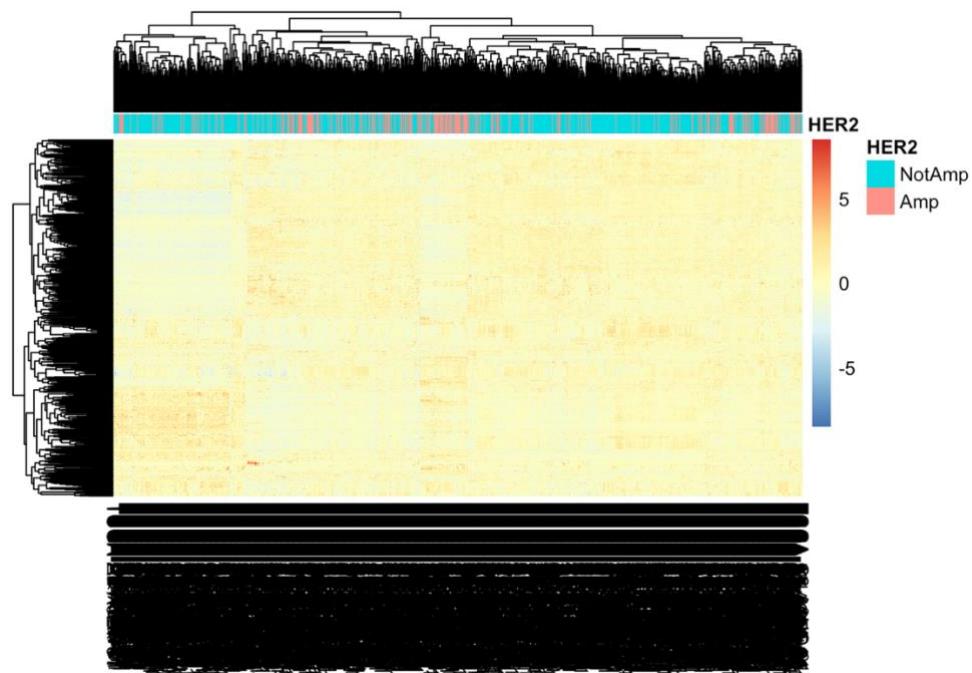
**Figure 3. PCA of HER2-amplified vs. non-amplified tumours.** Scatterplot of PC1 vs. PC2 using VST-normalised expression values. Each point: individual tumour sample.

Colour: HER2 status (Amp vs. NotAmp). PC1: 21% variance explained. PC2: 8% variance explained. The figure shows partial but not complete separation of the two groups along PC1, with substantial overlap across both components.

## 5. Heatmap of Top 500 Most Variable Genes

A heatmap was generated using the 500 genes with the highest variance across all tumour samples to examine broad expression patterns and sample-level clustering. Expression values were variance-stabilised and row-scaled to highlight relative up- and down-regulation across samples (Figure 4).

The heatmap reveals substantial transcriptional heterogeneity within both HER2-amplified and non-amplified groups. While some localised clusters of HER2-amplified samples are visible in the top annotation bar, HER2 status does not produce a dominant block-level clustering pattern, and samples from both groups are interspersed throughout the dendrogram. This indicates that although *ERBB2* amplification contributes to expression differences, it is not the primary driver of global clustering when considering the top variable genes. This is consistent with the known diversity of transcriptional subtypes within TCGA-BRCA.



**Figure 4. Heatmap of the 500 most variable VST-transformed genes.** Rows: individual genes. Columns: tumour samples. Colour scale: row-scaled expression values (blue = lower expression, yellow = higher expression). Top annotation: HER2 status (Amp vs. NotAmp). Dendograms: hierarchical clustering of samples and genes. The figure shows widespread variability in expression across tumours, with no single dominant cluster separating HER2-amplified from non-amplified samples.

## 6. LASSO Regularised Cox Regression

To assess whether expression patterns among the top differentially expressed genes were associated with overall survival, a LASSO-regularised Cox proportional hazards model was fitted using VST-transformed expression values of the top 100 DEGs. The *glmnet* model selected 12 genes with non-zero coefficients at the optimal  $\lambda$  value, indicating their contribution to the survival model (Table 2).

The selected genes included both positive and negative coefficients. Positive coefficients (e.g. *TEX19*, *ABCC2*, *CYB561*, *BRCA1*) indicate higher expression is associated with increased estimated hazard, whereas negative coefficients (e.g. *RPL19*, *RAPGEFL1*, *TMC3*, *CAPN6*) suggest a potential protective association. The magnitude of coefficients was modest, consistent with penalised regression shrinking effect sizes toward zero.

Gene	Coefficient
<i>STARD3</i>	0.010
<i>RPL19</i>	-0.109
<i>RAPGEFL1</i>	-0.175
<i>TEX19</i>	0.130
<i>THRA</i>	0.077
<i>MLLT6</i>	-0.050
<i>ABCC2</i>	0.128
<i>TMC3</i>	-0.118
<i>ZNF652</i>	-0.086
<i>CYB561</i>	0.110
<i>CAPN6</i>	-0.034
<i>BRCA1</i>	0.097

**Table 2. Genes with non-zero coefficients in the LASSO Cox survival model.** Positive values indicate an estimated increase in hazard with higher expression; negative values indicate an estimated decrease.

Overall, the results indicate that a subset of genes differentially expressed between HER2-amplified and non-amplified tumours carries prognostic information, although the effect sizes are small and heterogeneous. The presence of both ribosomal (e.g. *RPL19*) and signalling-related genes (e.g. *RAPGEFL1*) reflects the multifactorial relationships between transcriptional programs and clinical outcome in breast cancer.

## **Discussion**

The comparison of HER2-amplified and non-amplified breast tumours revealed widespread transcriptional alterations, with over 9,700 genes significantly differentially expressed. The most prominent pattern was a coordinated upregulation of mitotic cell-cycle pathways, accompanied by suppression of luminal differentiation markers and RNA-surveillance mechanisms. These findings directly support the rationale outlined in the background section, in which HER2 amplification was described as a driver of proliferative signalling and tumour aggressiveness (Cheng, 2024; Sorlie et al., 2001).

Focusing the biological interpretation on the cell-cycle/mitotic programme is justified for several reasons. First, this pathway family displayed the strongest enrichment in the Reactome analysis, with significant over-representation across G1/S transition, S phase, DNA synthesis, checkpoint control, and chromosomal segregation. Second, these processes align closely with established HER2 biology: HER2 amplification enhances signalling through PI3K/AKT and MAPK cascades, converging on E2F and MYC transcription factors that promote cyclin and CDK expression and activate DNA-replication machinery (Cheng, 2024). Experimental evidence has shown that HER2 signalling drives E2F-regulated DNA synthesis, whereas HER2 depletion reduces expression of cell-cycle genes and induces proliferative arrest (Nikolai et al., 2016). The enrichment of terms such as “Cell Cycle, Mitotic” and “Cell Cycle Checkpoints” therefore reflects a transcriptional profile that is biologically consistent with mechanisms described in the literature.

This proliferative signature also mirrors the clinical behaviour of HER2-positive tumours, which typically exhibit high proliferation indices and poor prognosis without targeted therapy (Cheng, 2024). In this study, the LASSO Cox model identified twelve differentially expressed genes with non-zero survival coefficients, indicating that components of the proliferative programme carry prognostic information beyond binary HER2 status. The presence of both positive and negative coefficients suggests functional diversity: some gene expression patterns may indicate highly proliferative and aggressive biology, whereas others may reflect compensatory or differentiation-linked states. More extensive modelling, incorporating treatment, intrinsic subtype, and stage, would be required to resolve these relationships fully.

The pathway analysis revealed an additional pattern: downregulation of mRNA surveillance and translation-quality pathways, including nonsense-mediated decay (NMD), termination, elongation, and ribosome-rescue processes. NMD prevents the accumulation of aberrant transcripts and can influence tumour evolution and immune presentation (Pan et al., 2025). Reduced activity of these pathways, observed here alongside strong mitotic activation, may permit tolerance of transcriptional errors during rapid proliferation, enhancing proteomic diversity and potentially contributing to genomic instability in HER2-amplified disease.

A further complementary pattern involved elevated expression of cancer-testis antigens (CTAs) and suppression of luminal differentiation markers (*CSN2*, *CSN3*, *LALBA*). CTAs are restricted to immune-privileged tissues such as testis but become aberrantly expressed in multiple cancers, where they have been explored as immunotherapy targets. Their induction in HER2-amplified tumours may reflect epigenetic remodelling associated with HER2-driven transformation and could mark biologically distinct or immunogenic subsets of HER2-positive disease. The suppression of luminal differentiation markers further indicates a shift away from a specialised mammary epithelial phenotype toward a more proliferative, less differentiated state.

The dominance of mitotic and cell-cycle pathways has therapeutic implications. The transcriptomic profile observed here provides a rationale for combining anti-HER2 therapies with inhibitors of CDK4/6 or other regulators of cell-cycle progression, an approach increasingly explored in HER2-positive breast cancer (Koirala et al., 2022). Moreover, HER2-amplified models show sensitivity to Aurora kinase inhibition, which induces mitotic catastrophe preferentially in tumours dependent on mitotic regulators (Tara et al., 2025). The enrichment of mitotic pathways in this dataset reinforces the biological rationale for these combination strategies.

Several limitations must be acknowledged. Bulk RNA-seq does not resolve cell-type heterogeneity or microenvironmental contributions. Clinical annotation was limited to overall survival, restricting prognostic modelling. Pathway enrichment analysis is correlative and dependent on curated gene-set definitions that may not capture all relevant biology. Further work integrating single-cell data, phosphoproteomics or functional perturbation experiments would be necessary to validate the dependency of HER2-amplified tumours on specific mitotic regulators identified here.

In summary, mitotic cell-cycle activation emerges as the dominant transcriptional hallmark of HER2-amplified breast tumours in this dataset. This signature integrates directly with the mechanistic background of HER2 signalling and is embedded within broader changes involving differentiation, RNA surveillance, and germline-like gene expression. These findings reinforce established models of HER2-driven tumour biology and highlight potential therapeutic vulnerabilities centred on proliferative machinery and translational control.

## Code Availability

All code used for data processing, differential expression analysis, pathway enrichment, visualisation, and survival modelling is available in the accompanying GitHub repository, [https://github.com/hannadmochowska/brc\\_a\\_erbb2/](https://github.com/hannadmochowska/brc_a_erbb2/).

## Reference list

- Baselga, J. and Swain, S.M. (2009). Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nature Reviews Cancer*, 9(7), pp.463–475. doi:<https://doi.org/10.1038/nrc2656>.
- cBioPortal (2018). *cBioPortal for Cancer Genomics*. [online] www.cbiportal.org. Available at: [https://www.cbiportal.org/study/summary?id=brca\\_tcga\\_pan\\_can\\_atlas\\_2018](https://www.cbiportal.org/study/summary?id=brca_tcga_pan_can_atlas_2018).
- Cheng, X. (2024). A Comprehensive Review of HER2 in Cancer Biology and Therapeutics. *Genes*, 15(7), pp.903–903. doi:<https://doi.org/10.3390/genes15070903>.
- Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1). doi:<https://doi.org/10.18637/jss.v033.i01>.
- Hynes, N.E. and Lane, H.A. (2005). ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Reviews Cancer*, 5(5), pp.341–354. doi:<https://doi.org/10.1038/nrc1609>.
- Koirala, N., Dey, N., Aske, J. and De, P. (2022). Targeting Cell Cycle Progression in HER2+ Breast Cancer: An Emerging Treatment Opportunity. *International Journal of Molecular Sciences*, 23(12), p.6547. doi:<https://doi.org/10.3390/ijms23126547>.
- Kolde, R. (2019). *pheatmap: Pretty Heatmaps*. [online] R-Packages. Available at: <https://cran.r-project.org/web/packages/pheatmap/index.html>.
- Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), p.550. doi:<https://doi.org/10.1186/s13059-014-0550-8>.
- Nikolai, B.C., Lanz, R.B., York, B., Dasgupta, S., Mitsiades, N., Creighton, C.J., Tsimelzon, A., Hilsenbeck, S.G., Lonard, D.M., Smith, C.L. and O’Malley, B.W. (2016). HER2 Signaling Drives DNA Anabolism and Proliferation through SRC-3 Phosphorylation and E2F1-Regulated Genes. *Cancer Research*, 76(6), pp.1463–1475. doi:<https://doi.org/10.1158/0008-5472.can-15-2383>.
- Pan, H., Wu, X., Hu, B., Xiong, H., Luo, Y. and Zhou, J.-K. (2025). Nonsense-mediated mRNA decay: a key regulatory system engaged in cancer. *Cell Communication and Signaling*, 23(1). doi:<https://doi.org/10.1186/s12964-025-02503-6>.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S.X., Lønning, P.E., Børresen-Dale, A.-L., Brown, P.O. and Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, [online] 406(6797), pp.747–752.  
doi:<https://doi.org/10.1038/35021093>.

Slamon, D., Clark, G., Wong, S., Levin, W., Ullrich, A. and McGuire, W. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, 235(4785), pp.177–182. doi:<https://doi.org/10.1126/science.3798106>.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Lonning, P.E. and Borresen-Dale, A.-L. . (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98(19), pp.10869–10874.  
doi:<https://doi.org/10.1073/pnas.191367098>.

Tara, S.A., Zekri, A., Sotoodehnejadnematalahi, F. and Modarressi, M.H. (2025). Aurora B inhibition induces polyploidy and mitotic catastrophe in HER2-amplified breast cancer: Telomere shortening as a potential anticancer mechanism of AZD1152-HQPA. *Biomedicine & Pharmacotherapy*, [online] 191, p.118509.  
doi:<https://doi.org/10.1016/j.biopha.2025.118509>.

Therneau, T. (2022). *A package for survival analysis in R*. [online] Available at: <https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf>.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. [online] Tidyverse.org. Available at: <https://ggplot2.tidyverse.org>.

Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., Tang, W., Wang, Q., Liu, B., Wang, R., Xie, W., Wu, T., Xie, L. and Yu, G. (2024). Using clusterProfiler to characterize multiomics data. *Nature Protocols*. [online] doi:<https://doi.org/10.1038/s41596-024-01020-z>.

Yu, G. and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Molecular BioSystems*, 12(2), pp.477–479.  
doi:<https://doi.org/10.1039/c5mb00663e>.