Subtype-Specific Transcriptional Dosage Buffering Predicts Disease-Specific Mortality in Breast Cancer

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Transcriptional dosage buffering—the resilience of gene expression to copy-number alterations—exhibits subtype-specific variation in breast cancer, but its impact on patient survival remains largely unexplored. To determine how localized buffering failures influence tumor lethality, we analyzed 1,417 primary breast tumors from the METABRIC dataset. We quantified the non-linear relationship between copy-number states and gene expression using a cross-validated gradient boosting framework, extracting directional buffering residuals that were subsequently aggregated into biological modules and adjusted for clinical covariates. We then investigated the role of promoter methylation in buffering discordance and applied Fine-Gray competing risks models to evaluate the prognostic value of these modules within specific Integrative Cluster (IntClust) subtypes. We identified two highly stable buffering networks, including an interferon-responsive module and an oncogenic regulator module. Interestingly, buffering failures were not associated with shifts in promoter methylation, pointing toward alternative regulatory mechanisms. Furthermore, intra-subtype survival analysis demonstrated that in IntClust 8 tumors, the inability to buffer interferon-responsive signaling independently predicted disease-specific mortality. Over-expression relative to copy-number expectations increased the risk of disease-specific death, while relative silencing conferred a protective effect. Ultimately, these results highlight non-linear transcriptional dosage buffering as a reproducible, subtype-specific molecular feature that offers prognostic insights independent of traditional clinical and genomic risk factors.

1 Introduction

Breast cancer is a highly heterogeneous disease characterized by extensive genomic instability. Copy-number alterations (CNAs) are primary drivers of this instability, frequently altering the transcriptomic landscape and promoting oncogenesis. However, the relationship between genomic dosage and transcriptional output is rarely strictly linear. Cells exhibit varying degrees of transcriptional dosage buffering—a phenomenon where gene expression remains resilient to underlying copy-number changes. While dosage buffering has been observed across various organisms and cancer types, its localized, subtype-specific rewiring in breast cancer remains poorly understood.

Furthermore, the clinical implications of buffering failures—specifically whether the inability to buffer distinct biological pathways contributes to tumor lethality—have not been systematically evaluated. Assessing this requires isolating disease-specific mortality from competing risks, such as age-related death, to ensure that buffering signatures reflect genuine tumor aggressiveness rather than systemic frailty or baseline genomic noise.

To address this gap, we investigated the non-linear coupling between copy-number states and gene expression across 1,417 primary breast tumors from the METABRIC cohort. We employed a cross-validated gradient boosting framework to model the expected expression from five-state CNA calls, extracting directional buffering residuals. These residuals capture two distinct modes of buffering failure: over-expression relative to genomic expectations (positive buffering discordance) and relative silencing (negative buffering discordance). By aggregating these gene-level metrics into established biological pathways and rigorously adjusting for clinical covariates—including non-synonymous tumor mutational burden, histological grade, age at diagnosis, and baseline proliferation—we identified stable, cohort-invariant buffering modules.

Our analysis reveals two highly stable buffering networks: an interferon-responsive module and an oncogenic regulator module. To understand the mechanistic basis of these buffering states, we investigated whether transcriptional discordance is driven by shifts in epigenetic regulation. Interestingly, buffering failures were not associated with changes in promoter methylation coupling, pointing toward alternative regulatory mechanisms such as distal enhancer activity or post-transcriptional feedback.

Finally, utilizing Fine-Gray competing risks models within specific Integrative Cluster (IntClust) subtypes, we demonstrate that localized buffering failures possess significant prognostic value. Specifically, in IntClust 8 tumors, the inability to buffer interferon-responsive signaling independently predicts disease-specific mortality. In this subtype, over-expression relative to copy-number expectations significantly increases the risk of disease-specific death, while relative silencing confers a protective effect. Ultimately, this study highlights non-linear transcriptional dosage buffering as a reproducible, subtype-specific molecular feature that offers prognostic insights orthogonal to traditional clinical and genomic risk factors.

2 Methods

2.1 Data acquisition and preprocessing

To investigate the localized rewiring of transcriptional dosage buffering, we utilized multi-omic data from the METABRIC cohort. We aligned clinical follow-up data, log2-transformed microarray gene expression, five-state copy-number alteration (CNA) calls, and reduced representation bisulfite sequencing (RRBS) promoter methylation beta values. The dataset was restricted to 1,417 primary breast tumor samples that were mutually present across all four molecular layers, as methylation profiling was the most restricted assay in the cohort. To focus the analysis on genes with dynamic transcriptional activity and to reduce the multiple-testing burden, we applied a variance filter to the expression matrix, retaining the top 5,000 most variable genes across the cohort.

2.2 Quantification of non-linear dosage buffering

We modeled the non-linear relationship between genomic dosage and transcriptional output using a gradient boosting regression framework. For each of the 5,000 retained genes, a LightGBM regressor was trained to predict log2 expression from the five-state CNA calls. To ensure that the captured dosage-expression discordance represents a stable property rather than model overfitting, we employed a 5-fold cross-validation strategy to compute out-of-fold predictions.

The buffering residual $r_{i,g}$ for patient i and gene g was defined as the difference between the observed expression $y_{i,g}$ and the model prediction $\hat{y}_{i,g}$. We decomposed these residuals into two directional components to capture distinct mechanisms of buffering failure: positive residuals representing over-expression relative to CNA expectations, and negative residuals representing relative silencing. These were calculated as:

$$r^{+}_{i,g} = \max(0, r_{i,g}) \tag{1}$$

$$r^{-}_{i,g} = |\min(0, r_{i,g})| \tag{2}$$

2.3 Pathway aggregation and covariate adjustment

Gene-level directional residuals were aggregated into biological modules using standard functional gene sets. Pathways were filtered to retain those with a minimum overlap of 15 genes with our 5,000-gene set and at least 70% coverage of their original size. For each patient and pathway, we calculated a positive buffering score and a negative buffering score by averaging the respective directional residuals of the constituent genes.

To ensure these scores were not merely proxies for genomic instability or tumor aggressiveness, we regressed out non-synonymous tumor mutational burden and histological grade using ordinary least squares regression. The residuals from these models served as the adjusted directional buffering scores. We further correlated these adjusted scores with age at diagnosis and the THREEGENE proliferation score to monitor for confounding by systemic frailty or baseline proliferation. Finally, to confirm that the buffering landscape is a reproducible biological property rather than a technical artifact, we evaluated the stability of the

Score Type	n genes	Median ρ (Concordant)	Median ρ (Discordant)	MWU p-value
'mod_15_pos'	21	-0.0434	-0.0341	0.500
'mod_15_neg'	21	-0.0475	-0.0438	0.520
'mod_27_neg'	95	-0.0418	-0.0401	0.604

adjusted scores across the five METABRIC accrual cohorts using a Kruskal-Wallis H-test, retaining only pathways with statistically invariant distributions ($p > 0.05$).

2.4 Epigenetic coupling analysis

To determine whether buffering failures were driven by a loss of epigenetic regulation, we cross-referenced the discordant pathways with promoter methylation data. Patients were grouped by their Integrative Cluster (IntClust) designation. Within each group, tumors in the top and bottom 20% of the adjusted buffering score distributions were classified as "discordant," while the middle 60% were classified as "concordant."

For genes within the highly stable pathways (such as the interferon-responsive Module 15 and the oncogenic Module 27), we computed the Spearman correlation (ρ) between methylation beta values and log2 expression, utilizing pairwise deletion to gracefully handle missing methylation data. We then compared the methylation-expression coupling strength between discordant and concordant tumors using a Mann-Whitney U (MWU) test to evaluate shifts in epigenetic constraint (Table 2.4).

2.5 Intra-subtype competing risks survival analysis

The prognostic utility of the stable directional buffering scores was evaluated strictly within individual IntClust groups to account for baseline subtype-specific risks. We excluded IntClust groups with fewer than 15 disease-specific deaths to ensure statistical robustness and avoid spurious hazard ratios.

For each valid IntClust group and stable pathway, we fitted a Fine-Gray competing risks model. The event of interest was explicitly defined as "Died of Disease," with "Died of Other Causes" treated as a competing event. The adjusted positive and negative buffering scores were used as primary predictors. To validate that the buffering mechanism is an independent axis of lethality, the models were adjusted for age at diagnosis, tumor size, number of positive lymph nodes, and the THREEGENE proliferation score. The proportional subdistribution hazards assumption was verified for all models. Finally, we applied a Benjamini-Hochberg False Discovery Rate (FDR) correction to the hazard ratio p-values within each IntClust group to account for multiple testing across pathways, extracting the subdistribution hazard ratios (sHR), 95% confidence intervals, and adjusted p-values.

3 Results

3.1 Identification of stable transcriptional dosage buffering modules

To map the landscape of transcriptional dosage buffering across breast cancer subtypes, we analyzed 1,417 primary tumors from the METABRIC cohort with matched expression, copy-number, and methylation profiles. Using our cross-validated gradient boosting framework, we quantified the non-linear coupling between genomic dosage and transcriptional output, extracting positive and negative buffering scores for established biological pathways. These scores represent the magnitude of over-expression and relative silencing, respectively, compared to copy-number expectations.

To ensure that these buffering signatures reflect specific transcriptional rewiring rather than broad genomic instability or systemic frailty, the scores were adjusted for non-synonymous tumor mutational burden, histological grade, age at diagnosis, and baseline proliferation. Stability testing across the five accrual cohorts identified two highly reproducible buffering networks (Kruskal-Wallis $p > 0.05$): Module 15, which is strongly enriched for interferon-responsive genes (e.g., *IFI6*, *ISG15*, *STAT1*), and Module 27, which comprises oncogenic regulators (e.g., *ERBB2*, *S100P*).

Visualizing these adjusted directional buffering scores revealed a complex, localized landscape of transcriptional resilience (Figure 1). While specific buffering states partially align with established Integrative



Figure 1: This heatmap displays the clustered landscape of stable directional buffering scores, quantifying transcriptional resilience to copy-number alterations across breast tumour samples. Rows represent adjusted residuals for the positive and negative buffering of Module 15 (`mod_15_pos`, `mod_15_neg`) and negative buffering of Module 27 (`mod_27_neg`), while columns represent individual samples (`SAMPLE_ID`) annotated by Integrative Cluster, subtype, and accrual cohort. The clustering illustrates that although specific buffering states align with established molecular subtypes, significant intra-subtype heterogeneity remains. By adjusting these scores for tumour mutational burden, histological grade, age, and proliferation, the plot reveals a localized, non-linear axis of tumour organization that is independent of systemic genomic instability and established clinical taxonomies.

Cluster (IntClust) and claudin-low subtypes, substantial intra-subtype heterogeneity remains. This indicates that dosage buffering constitutes a distinct, non-linear axis of tumor organization that is not fully captured by traditional clinical or genomic taxonomies.

3.2 Buffering discordance is independent of promoter methylation

We next investigated the mechanistic basis of buffering failure, specifically testing the hypothesis that the transition from a buffering-concordant to a discordant state is driven by a loss of epigenetic regulation. We compared the Spearman correlation (ρ) between promoter methylation and gene expression for tumors exhibiting extreme buffering discordance (the top and bottom 20% of the score distribution) against those maintaining concordant buffering (the middle 60%).

Our analysis revealed a null shift in epigenetic constraint across all stable modules (Table 2.4). The strength of methylation-expression coupling was statistically indistinguishable between buffering-concordant and discordant tumors for both the positive and negative buffering states of Module 15, as well as the negative buffering state of Module 27 (Figure 2). These results demonstrate that the observed non-linear dosage buffering is not primarily mediated by promoter-level methylation shifts, suggesting the involvement

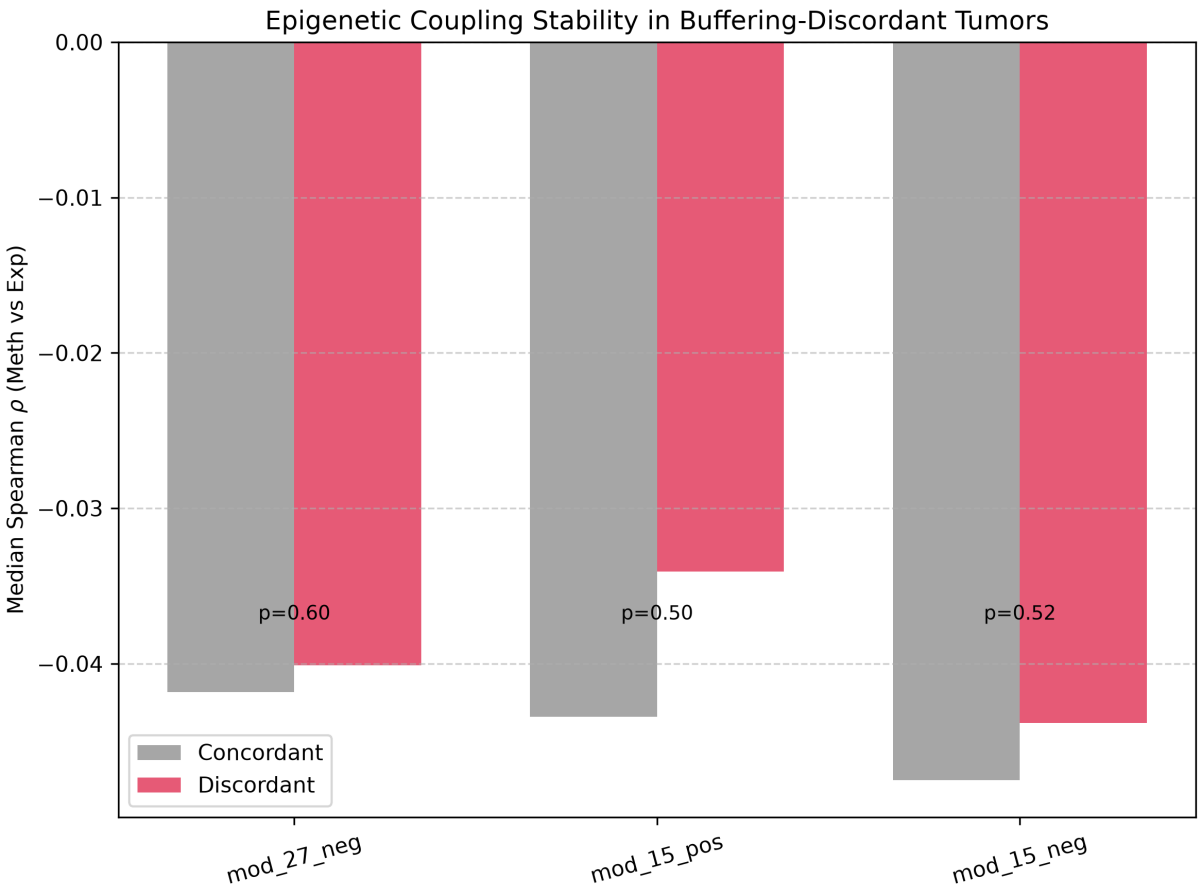


Figure 2: Comparison of the median Spearman correlation (ρ) between promoter methylation and gene expression for buffering-concordant (middle 60% of the distribution) and discordant (top/bottom 20%) tumours across three stable module scores (`mod_27_neg`, `mod_15_pos`, and `mod_15_neg`). The lack of significant difference in epigenetic coupling between the two groups demonstrates that the failure of transcriptional dosage buffering is not driven by a loss of promoter methylation regulation. This suggests that the observed non-linear dosage buffering operates independently of promoter-level epigenetic shifts and may instead depend on distal regulatory elements or post-transcriptional feedback.

of alternative regulatory mechanisms such as distal enhancer activity or post-transcriptional feedback.

3.3 Subtype-specific prognostic impact of buffering failures

To determine whether localized buffering failures influence tumor lethality, we evaluated the prognostic impact of the stable buffering modules using Cause-Specific Hazard (CSH) models. By strictly stratifying patients by their IntClust designation and adjusting for age, tumor size, lymph node status, and baseline proliferation, we isolated the independent effect of buffering on disease-specific mortality from competing risks such as age-related death.

The most significant prognostic signal was observed within the IntClust 8 subgroup ($n = 230$, including 61 disease-specific deaths). In these tumors, the inability to buffer interferon-responsive signaling (Module 15) emerged as a strong, independent predictor of disease-specific mortality. Specifically, positive buffering discordance (`mod_15_pos`)—indicating higher over-expression relative to copy-number expectations—was associated with a significantly increased risk of disease-specific death (sHR = 2.827, 95% CI: 1.419–5.632, FDR = 0.0055). Conversely, negative buffering discordance (`mod_15_neg`), or relative silencing of this module,

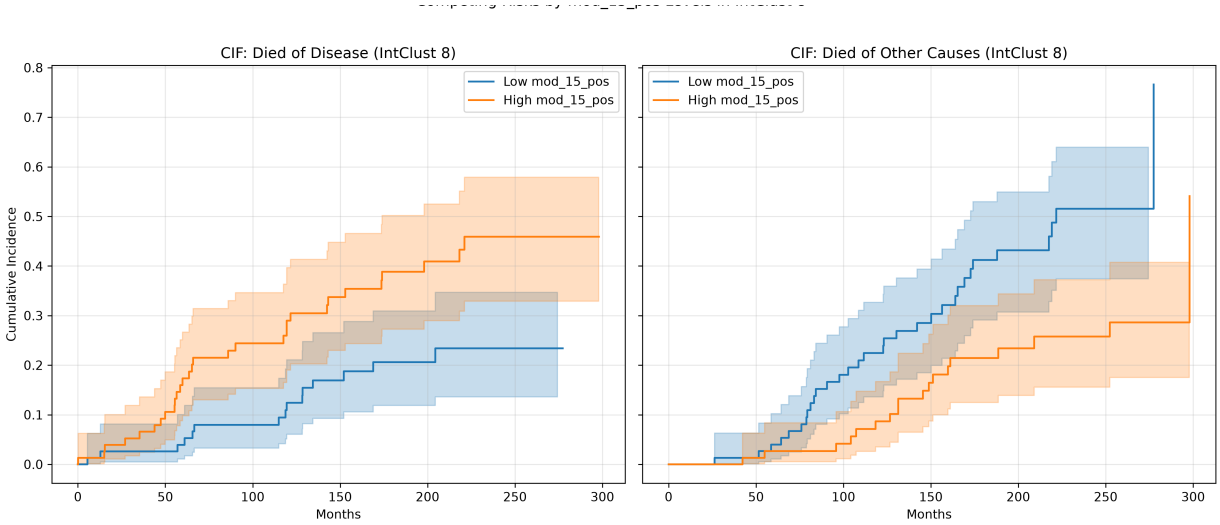


Figure 3: Cumulative Incidence Functions for patients in the IntClust 8 subgroup stratified by positive transcriptional dosage buffering of Module 15 (`mod_15_pos`). The left panel demonstrates that high `mod_15_pos` scores, which indicate the over-expression of interferon-responsive genes relative to copy-number alterations, are associated with a significantly increased risk of disease-specific mortality. Conversely, the right panel shows that the incidence of death from other causes remains uniform across these buffering groups. This indicates that the failure to buffer interferon-responsive signaling is an independent predictor of breast cancer mortality, specifically stratifying patients by tumour-specific lethality rather than systemic frailty or competing risks.

conferred a protective effect and was associated with improved outcomes (sHR = 0.281, 95% CI: 0.120–0.662, FDR = 0.0055).

Cumulative Incidence Functions for IntClust 8 patients illustrate that high Module 15 positive buffering scores specifically stratify patients by death from disease, whereas the incidence of death from other causes remains uniform across the buffering groups (Figure 3). This confirms that the buffering failure reflects genuine tumor-specific lethality rather than systemic frailty.

Furthermore, a forest plot of hazard ratios across all evaluated Integrative Clusters confirms that the prognostic value of dosage buffering is highly subtype-specific (Figure 4). While other subgroups exhibited observable trends—such as increased risk associated with `mod_15_pos` in IntClust 1 (sHR = 1.929, $p = 0.0965$) and `mod_27_neg` in IntClust 4ER- (sHR = 4.661, $p = 0.2087$)—these effects did not reach statistical significance after multiple-testing correction. Thus, the failure to buffer interferon-responsive signaling represents a highly localized, independent axis of lethality specifically within IntClust 8 breast tumors.

4 Conclusions

In this study, we systematically evaluated the landscape and clinical implications of transcriptional dosage buffering in breast cancer. By integrating copy-number, expression, and methylation profiles across 1,417 primary tumors from the METABRIC cohort, we demonstrated that the non-linear resilience of gene expression to genomic dosage is a stable, reproducible molecular feature. Rather than being a byproduct of broad genomic instability or systemic frailty, localized buffering failures represent a distinct axis of tumor organization.

Our framework identified two highly stable buffering networks, most notably an interferon-responsive module and an oncogenic regulator module. Mechanistically, we established that the failure to buffer these pathways is not driven by a loss of promoter methylation regulation. The preservation of epigenetic coupling in highly discordant tumors suggests that non-linear dosage buffering relies on alternative regulatory layers, such as distal enhancer modulation, chromatin architecture, or post-transcriptional feedback mechanisms.

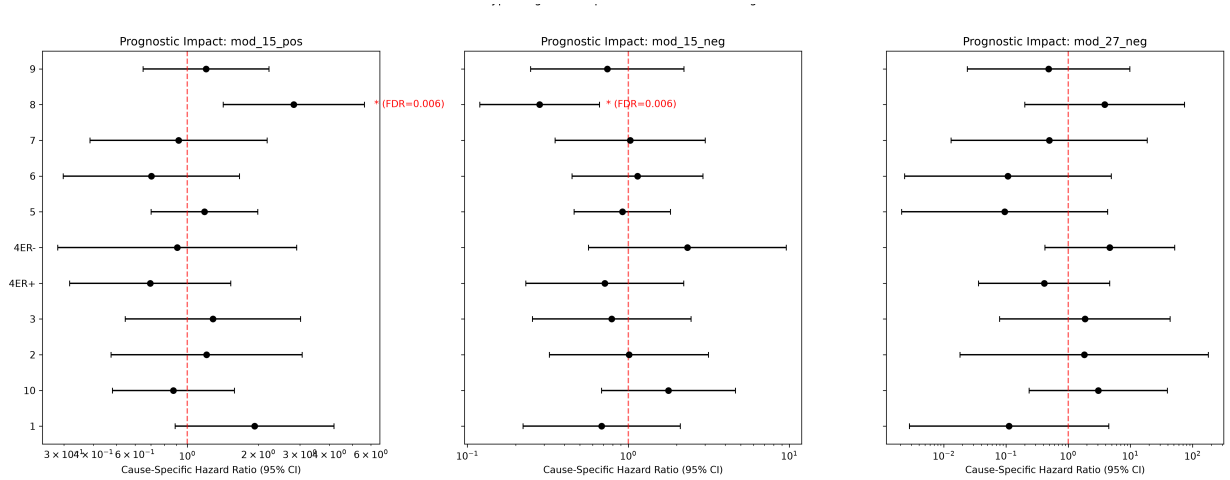


Figure 4: Forest plot of Cause-Specific Hazard Ratios (with 95% confidence intervals) illustrating the subtype-specific prognostic impact of transcriptional dosage buffering across Integrative Clusters (`INTCLUST`). The panels display the independent effect of positive and negative buffering scores for Modules 15 and 27 on disease-specific mortality, adjusted for clinical covariates such as age, tumour size, lymph node status, and baseline proliferation. The plot demonstrates that the prognostic value of dosage buffering is highly localized to specific breast cancer subtypes. Most notably, in IntClust 8, the failure to buffer interferon-responsive genes (Module 15) is a strong predictor of mortality: positive buffering (`mod_15_pos`) significantly increases disease-specific risk, while negative buffering (`mod_15_neg`) improves outcomes (FDR = 0.006). While other groups like IntClust 1 and 4ER- show observable trends for `mod_15_pos` and `mod_27_neg` respectively, they do not reach statistical significance, highlighting that this non-linear axis of tumour organization provides distinct, subtype-specific prognostic information.

Crucially, we revealed that the prognostic impact of buffering discordance is highly subtype-specific. By utilizing competing risks models to isolate disease-specific mortality, we found that the inability to buffer interferon-responsive signaling is a strong, independent predictor of lethality specifically within the IntClust 8 subgroup. In these tumors, over-expression relative to copy-number expectations significantly exacerbates the risk of disease-specific death, while relative silencing provides a protective effect.

Ultimately, these findings highlight that transcriptional dosage buffering provides critical prognostic insights orthogonal to established clinical covariates, baseline proliferation, and traditional genomic taxonomies. Recognizing and quantifying these non-linear multi-omic interactions offers a refined understanding of subtype-specific tumor aggressiveness and presents a novel avenue for improving patient stratification in breast cancer.