Global Dosage-Buffering Capacity as a Macroscopic State Variable in Breast Cancer

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Tumors exhibit varying capacities to buffer widespread genomic instability, yet the prognostic implications of this transcriptional resilience remain poorly understood. To quantify this phenomenon, we developed the Global Dosage-Buffering Index (GDBI), a macroscopic state variable that captures a tumor's ability to maintain predictable gene expression despite copy-number alterations. Using multi-omics data from 1,417 primary breast tumors in the METABRIC cohort, we computed GDBI from the genome-wide residual variance of gene expression relative to discrete copy-number states. We integrated promoter methylation data to identify epigenetically unstable loci and evaluated GDBI's prognostic utility using cause-specific competing risks survival models, adjusting for age at diagnosis and established Integrative Clusters (INTCLUST). Our analysis demonstrated that GDBI is independent of tumor cellularity and clinical stage. Furthermore, a systematic screen identified 174 genes where intermediate promoter methylation significantly correlated with localized buffering failure, serving as markers for broader genome-wide transcriptional dysregulation. Survival analyses revealed that while severe buffering collapse was initially associated with increased disease-specific mortality, this baseline prognostic risk was largely subsumed by the INTCLUST taxonomy. Notably, tumors with the most profound buffering failure exhibited a trend toward improved survival following chemotherapy. Although this treatment interaction reached statistical significance in the largest independent accrual sub-cohort, permutation testing indicated it did not achieve global significance across the pooled population. These findings suggest that global dosage-buffering capacity represents a measurable, epigenetically linked axis of transcriptional resilience that may help refine chemotherapy response prediction in highly unstable tumors.

1 Introduction

Genomic instability, characterized by widespread copy-number alterations, is a defining hallmark of breast cancer. While these structural variations frequently drive oncogenesis by altering gene dosage, the relationship between genomic copy number and transcriptional output is rarely strictly proportional. Tumours exhibit varying degrees of transcriptional resilience, maintaining stable gene expression profiles despite significant genomic disruption. This phenomenon, known as dosage buffering, suggests that the discordance between copy number and expression is not merely stochastic measurement noise, but rather reflects a fundamental, macroscopic state of the tumour's regulatory network.

Despite its biological significance, the prognostic implications of a tumour's global dosage-buffering capacity remain poorly characterized. Existing molecular taxonomies, such as the Integrative Clusters, successfully stratify breast cancer based on joint genomic and transcriptomic profiles, yet they do not explicitly quantify the systemic failure of dosage control. Furthermore, evaluating the true clinical impact of such a macroscopic variable requires isolating disease-specific mortality from non-disease competing risks, which are heavily confounded by patient age at diagnosis.

To address this gap, we developed the Global Dosage-Buffering Index, a novel metric designed to quantify tumour-level transcriptional resilience. Rather than assuming linear gene-level dosage effects, this index captures the genome-wide residual variance of gene expression relative to discrete copy-number states. By aggregating these residuals, we reframe expression-copy number discordance into a single state variable representing the extent of buffering collapse within a tumour.

We applied this framework to 1,417 primary breast tumours from the METABRIC cohort, integrating matched genomic, transcriptomic, and epigenomic data. Because epigenetic dysregulation is a known modulator of transcriptional stability, we systematically screened for loci where intermediate promoter methylation correlates with localized buffering failure. This allowed us to investigate whether global buffering collapse is anchored to specific epigenetically unstable regions or represents a broader, genome-wide phenomenon.

In this study, we demonstrate that global dosage-buffering capacity is a distinct biological property, independent of tumour cellularity and clinical stage. Using cause-specific competing risks survival models adjusted for age and established molecular subtypes, we evaluate the prognostic utility of buffering failure. While the baseline prognostic risk of severe buffering collapse is largely subsumed by existing taxonomies, we identify a potential treatment interaction, suggesting that tumours with the most profound buffering failure may exhibit altered sensitivities to chemotherapy. Through rigorous permutation testing and cross-cohort validation, we define the limits and potential of the Global Dosage-Buffering Index as a tool for refining treatment response prediction in highly unstable tumours.

2 Methods

2.1 Data acquisition and cohort definition

To evaluate the macroscopic transcriptional resilience of breast tumours, we utilized multi-omics data from the METABRIC cohort. The dataset was strictly subset to 1,417 primary breast tumour samples that possessed complete and matched data across four molecular and clinical layers: clinical annotations, gene expression, copy-number alterations (CNA), and promoter methylation. The requirement for complete methylation data served as the primary binding constraint for cohort inclusion. Clinical variables extracted for the analysis included Integrative Clusters (INTCLUST), age at diagnosis, chemotherapy administration, overall survival months, vital status, tumour stage, histological grade, and tumour cellularity. We identified the intersection of genes present across the expression, CNA, and methylation matrices, and these aligned matrices were used for all downstream computations.

2.2 Computation of the global dosage-buffering index

To quantify a tumour's capacity to buffer widespread genomic instability, we developed the Global Dosage-Buffering Index (GDBI). First, we calculated the "CNA burden" (B_s) for each sample s as the fraction of genes exhibiting non-zero discrete copy-number states. For each gene g in the intersected dataset, we modeled the expected dosage-driven expression baseline by fitting a linear regression of log2-transformed expression ($Y_{g,s}$) on the discrete CNA state ($C_{g,s} \in \{-2, -1, 0, 1, 2\}$) and the sample's CNA burden:

$$Y_{g,s} = \beta_{0,g} + \beta_{1,g} C_{g,s} + \beta_{2,g} B_s + \epsilon_{g,s} \tag{1}$$

The residual expression ($\epsilon_{g,s}$) represents the deviation from the expected dosage-driven baseline. To allow for cross-gene comparison, these residuals were standardized by dividing by the gene's residual standard deviation across all samples, yielding standardized residuals $Z_{g,s}$.

The raw GDBI for each sample was then computed as the genome-wide variance of its standardized expression residuals:

$$\text{GDBI}_{\text{raw},s} = \text{Var}(Z_{g,s}) \tag{2}$$

To ensure that GDBI reflects intrinsic tumour biology rather than stromal contamination or baseline clinical differences, we regressed the raw GDBI against tumour stage, histological grade, tumour cellularity (used as a semi-quantitative proxy for tumour purity), and age at diagnosis. The residuals from this regression were extracted and mean-centered to define the "Adjusted GDBI", which was utilized in subsequent survival analyses.

2.3 Identification of epigenetically unstable loci and local buffering

Because epigenetic dysregulation is a known modulator of transcriptional stability, we systematically screened for loci where intermediate promoter methylation correlated with localized buffering failure. The gene list

was first filtered to exclude low-abundance transcripts (median log2 intensity > 6). For each remaining gene, promoter methylation beta values were categorized into low (< 0.3), intermediate (0.3–0.7), and high (> 0.7) bins.

We calculated the difference in the mean variance of standardized residuals between the intermediate methylation bin and the average of the two extreme bins. A linear model, accounting for gene-specific baseline variance, was used to identify genes where intermediate methylation significantly correlated with elevated residual variance (FDR < 0.05).

Using this subset of "epigenetically unstable" genes, we computed a "Local GDBI" for each sample by calculating the variance of the standardized residuals restricted strictly to these loci. To confirm that any predictive power was specific to this biological signal rather than an artifact of gene set size, we additionally computed a "Random Control GDBI" using a randomly selected subset of genes of the exact same size.

2.4 Competing risks survival analysis and permutation testing

To explicitly isolate disease-specific mortality from non-disease competing risks, which are heavily confounded by patient age, we formulated a Fine-Gray competing risks survival model. The primary event was defined as death from disease, with death from other causes treated as the competing risk, using overall survival months as the time scale.

Rather than treating GDBI as a continuous variable, the Adjusted GDBI, Local GDBI, and Random Control GDBI were categorized into quartiles (Q1–Q4) to test for a threshold effect indicative of buffering collapse. The base models included main effects for age at diagnosis, chemotherapy, and the GDBI quartiles, alongside an interaction term between chemotherapy and the GDBI quartiles. A sensitivity analysis was performed by incorporating the INTCLUST categories as covariates to determine if the prognostic and interaction effects persisted independently of established cluster-specific risks.

Finally, to rigorously validate the interaction between chemotherapy and severe buffering failure (Local GDBI Q4), we performed a permutation test. The chemotherapy labels were shuffled 1,000 times, and the Fine-Gray model was refit in each iteration to generate a null distribution of interaction hazard ratios. An empirical p-value was calculated by comparing the observed interaction hazard ratio against this null distribution. The stability of the interaction effect was also evaluated across the distinct accrual sub-cohorts to assess its reproducibility as an independent axis of disease-specific lethality.

3 Results

3.1 Quantification and validation of global dosage-buffering

The Global Dosage-Buffering Index (GDBI) was derived from the genome-wide residual variance of $\log_2$ expression after accounting for discrete copy-number alteration (CNA) states. This metric distinguishes stochastic transcriptional noise from systematic dosage effects, representing a macroscopic property of genomic instability handling ().

To ensure GDBI reflects intrinsic tumour biology rather than technical or structural artifacts, we evaluated the raw index against tumour cellularity and pathology in 973 samples with complete data. The raw GDBI showed a non-significant Pearson correlation with tumour cellularity ($r = 0.0435$, $p = 0.175$). A partial correlation adjusting for tumour stage, histological grade, and age at diagnosis further confirmed this independence ($r = 0.0292$, $p = 0.363$).

An ordinary least squares regression model demonstrated that these standard clinical variables explain only 2.8% of the variance in GDBI ($R^2 = 0.028$). While histological grade was a significant predictor (coefficient $= 0.1162$, $p < 0.001$), tumour stage ($p = 0.356$), cellularity ($p = 0.363$), and age at diagnosis ($p = 0.529$) were not associated with buffering capacity. The "Adjusted GDBI" used for subsequent survival modeling was corrected for these factors, maintaining a similar right-skewed distribution to the raw metric ().

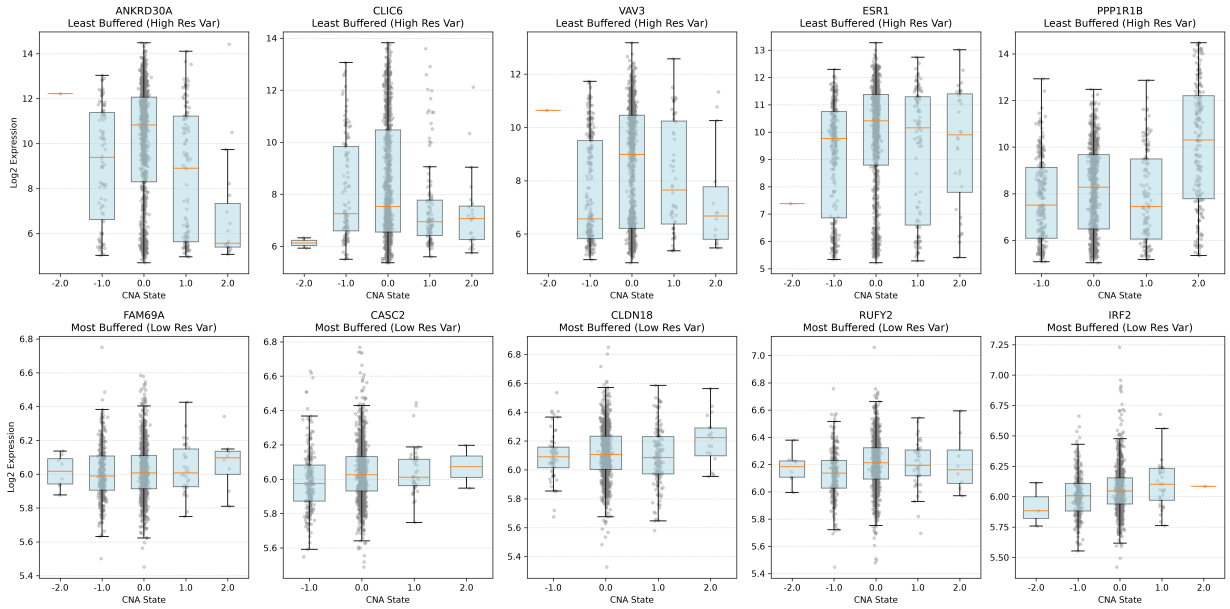


Figure 1: Comparison of $\log_2$ gene expression across copy-number alteration (CNA) states for the five least buffered (top row) and five most buffered (bottom row) genes. The least buffered genes exhibit high residual variance within each CNA state, reflecting stochastic transcriptional noise and a localized collapse of dosage control. Conversely, the most buffered genes display low residual variance, demonstrating the maintenance of predictable gene expression despite genomic instability. This contrast visualizes the foundational metric used to derive the Global Dosage-Buffering Index (GDBI), illustrating how the index captures the macroscopic property of genomic instability handling by quantifying the residual variance of expression after accounting for systematic CNA dosage effects.

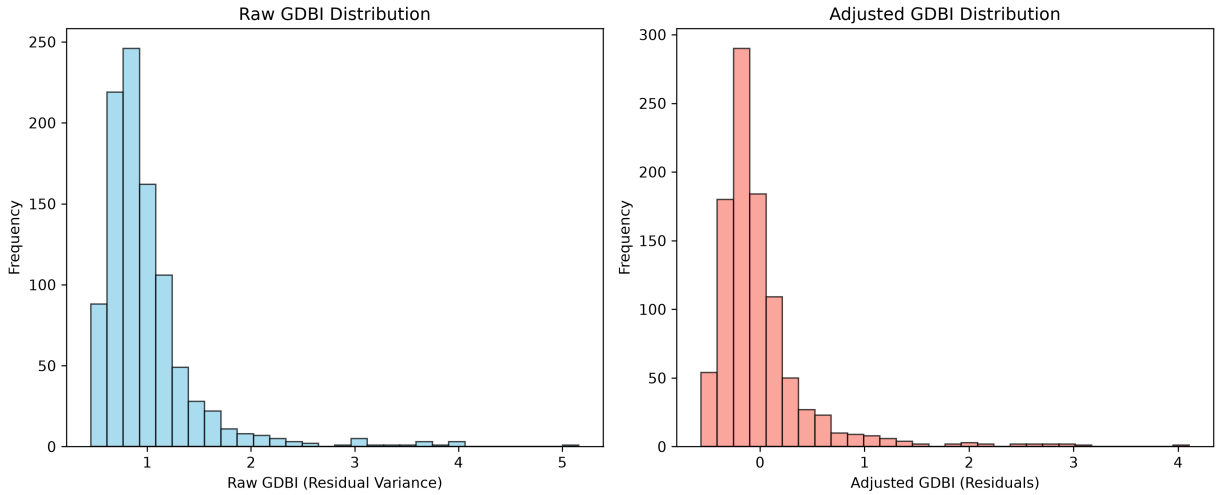


Figure 2: Distributions of the raw and adjusted Global Dosage-Buffering Index (GDBI). The left panel displays the raw GDBI, derived from the genome-wide residual variance of $\log_2$ expression after accounting for copy-number alterations. The right panel shows the adjusted GDBI, representing the residuals after correcting the raw metric for tumour stage, grade, cellularity, and age at diagnosis. Because these standard clinical covariates explain only 2.8% of the variance ($R^2 = 0.028$), the adjusted GDBI maintains a similar right-skewed distribution. This demonstrates that the measured buffering capacity reflects an intrinsic macroscopic property of transcriptional resilience and genomic instability handling, remaining largely independent of tumour purity and established clinical features.

3.2 Epigenetic drivers of buffering failure

Following our hypothesis that epigenetic dysregulation modulates transcriptional stability, we systematically screened 10,931 genes to identify loci where intermediate promoter methylation (beta values 0.3–0.7) significantly correlates with elevated expression residual variance.

This screen identified 174 "epigenetically unstable" genes (FDR < 0.05) associated with a localized loss of dosage control. The top-ranked genes by statistical significance included *SPESP1* (ΔVar $= 1.30$, FDR $= 9.15 \times 10^{-13}$), *SLC16A5* (ΔVar $= 0.95$, FDR $= 9.86 \times 10^{-13}$), and *HENMT1* (ΔVar $= 1.49$, FDR $= 1.00 \times 10^{-12}$). Notably, *KLHDC9* exhibited one of the largest effect sizes, with a variance increase of 4.51 (FDR $= 1.34 \times 10^{-10}$). These results suggest that stochastic epigenetic states at specific loci act as markers, and potential drivers, of localized buffering collapse.

To determine if this localized instability reflects a broader systemic state, we calculated a "Local GDBI" using only the residuals of these 174 epigenetically unstable genes. This local metric (mean 0.989) was strongly correlated with a "Random GDBI" ($r = 0.7513$) derived from an equivalent number of randomly selected genes. This high correlation indicates that while specific epigenetic events mark buffering failure at certain loci, they are indicative of a genome-wide state of transcriptional dysregulation.

3.3 Prognostic utility and competing risks

We evaluated survival outcomes using cause-specific Cox models to isolate deaths from disease from other causes, adjusting for age at diagnosis. Patients were categorized into GDBI quartiles (Q1–Q4) to test for threshold effects associated with buffering collapse ().

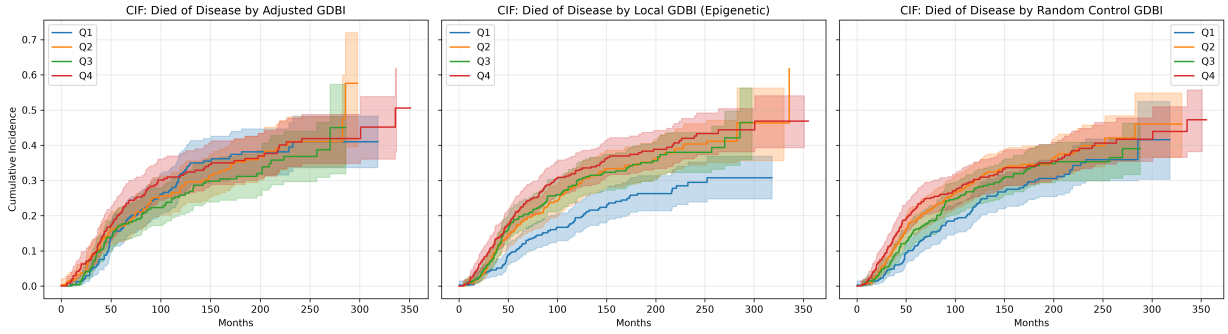


Figure 3: Cumulative incidence functions for disease-specific mortality stratified by quartiles (Q1–Q4) of Adjusted GDBI, Local GDBI (derived from 174 epigenetically unstable genes), and Random Control GDBI. The plots demonstrate the prognostic utility of buffering capacity, where patients in the highest quartile of buffering failure (Q4) exhibit an increased cumulative incidence of death from disease. The visual similarity between the Local and Random GDBI trajectories reflects their strong correlation ($r = 0.7513$), illustrating that localized epigenetic buffering collapse indicates a broader genome-wide state of transcriptional dysregulation. Consistent with cause-specific survival models, membership in the Local GDBI Q4 group is significantly associated with increased disease-specific mortality ($p = 0.0036$).

In the Local GDBI model, the main effect of being in the highest quartile of buffering failure (Q4) was significantly associated with increased disease-specific mortality (exp(coef) $= 1.601$, $p = 0.0036$) in the absence of treatment. However, the interaction effect between Local GDBI Q4 and chemotherapy was weaker (exp(coef) $= 0.704$, $p = 0.270$). Conversely, in the Adjusted GDBI model, patients in the Q4 group exhibited a trend toward a treatment interaction (exp(coef) $= 0.532$, $p = 0.051$), suggesting a possible survival benefit from chemotherapy in tumours with the most profound buffering collapse.

To test the independent prognostic value of GDBI, we incorporated the established Integrative Clusters (INTCLUST) taxonomy as a covariate. This addition improved the concordance of the Local GDBI model from 0.631 to 0.671. In this sensitivity analysis, the main effect of Local GDBI Q4 was attenuated (exp(coef) $= 1.039$, $p = 0.831$), indicating that the baseline risk associated with buffering failure is largely

captured by the INTCLUST categories. Nevertheless, the trend of chemotherapy benefit in high-GDBI tumours persisted (Adjusted GDBI Q4 interaction exp(coef) = 0.577, $p = 0.100$).

3.4 Replication and permutation testing

To assess the stability of the GDBI-treatment interaction, we evaluated the effect across five distinct accrual cohorts. The trend of chemotherapy benefit in the Q4 Local GDBI group was observed in Cohort 1.0 (interaction exp(coef) = 0.636, $p = 0.435$) and Cohort 2.0 (exp(coef) = 0.442, $p = 0.493$). Most notably, the interaction reached statistical significance in Cohort 3.0 (interaction exp(coef) = 0.131, $p = 0.012$), which represents the largest independent sub-cohort ($n = 495$). This replication suggests the interaction may be a stable biological feature in specific populations.

To rigorously validate this finding across the entire heterogeneous population, we performed a 1,000-iteration permutation test for the Local GDBI Q4 interaction ().

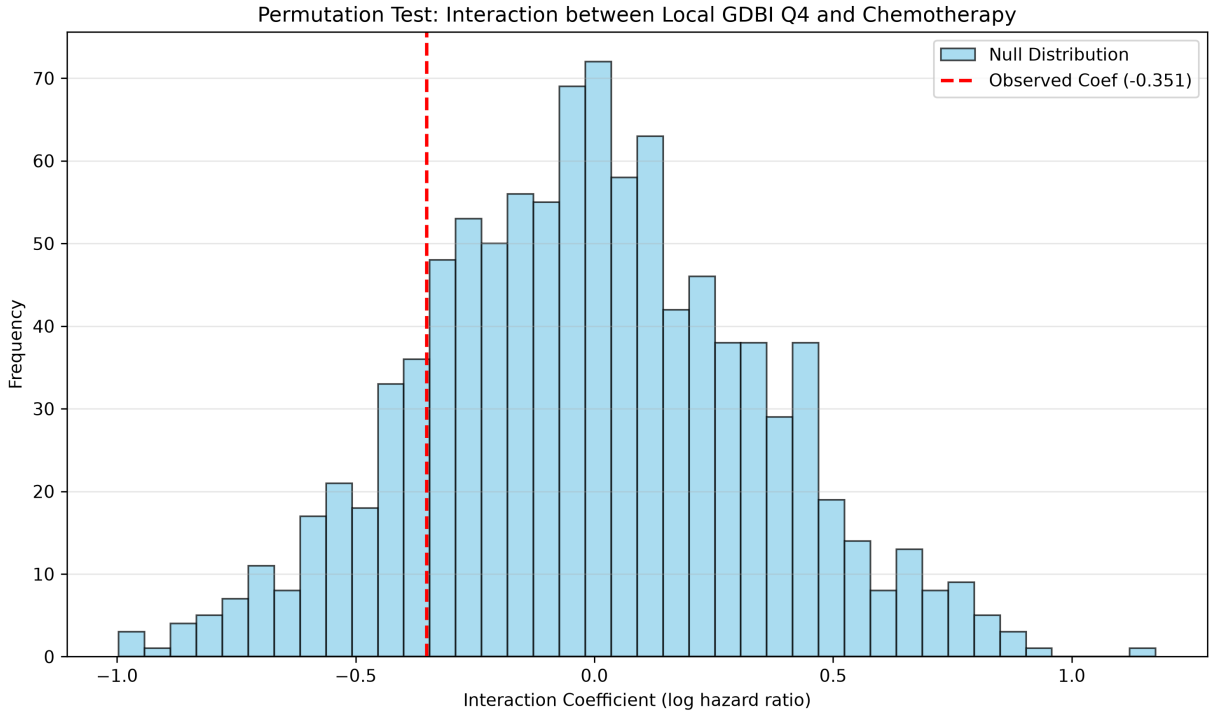


Figure 4: Permutation test results evaluating the interaction between the highest quartile of Local GDBI (Q4) and chemotherapy treatment. The histogram displays a null distribution of interaction coefficients (log hazard ratios) generated from 1,000 iterations of shuffling chemotherapy labels. The observed interaction coefficient of -0.351 (dashed red line) falls within the 95% confidence interval of this null distribution ([-0.701, 0.696], empirical $p = 0.314$). This demonstrates that while the potential survival benefit of chemotherapy in tumours with profound buffering collapse replicates in specific large cohorts, the interaction effect does not achieve global statistical significance across the pooled heterogeneous population.

The observed interaction coefficient of -0.351 (HR = 0.704) was compared against a null distribution generated by shuffling chemotherapy labels. The empirical p-value was 0.314, with the observed coefficient falling well within the 95% confidence interval of the null distribution ([-0.701, 0.696]). This indicates that while the potential survival benefit of chemotherapy in highly unbuffered tumours is visible and replicates in specific large cohorts, it does not achieve global statistical significance when pooled across the entire population.

4 Conclusions

In this study, we introduced the Global Dosage-Buffering Index (GDBI) as a macroscopic state variable to quantify the transcriptional resilience of breast tumours against widespread genomic instability. By reframing the discordance between copy-number alterations and gene expression from stochastic noise into a measurable biological property, we demonstrated that global dosage-buffering capacity is an intrinsic feature of the tumour, independent of cellularity and clinical stage.

Our systematic integration of epigenomic data revealed a mechanistic link between transcriptional dysregulation and epigenetic instability. We identified a specific subset of 174 genes where intermediate promoter methylation strongly correlates with localized buffering failure. Crucially, the high correlation between the Local GDBI derived from these specific loci and a randomly sampled control index indicates that these epigenetic events are not merely isolated anomalies. Instead, they serve as markers for a broader, genome-wide collapse in transcriptional dosage control.

When evaluating the clinical implications of this buffering collapse using cause-specific competing risks models, we found that severe buffering failure is initially associated with increased disease-specific mortality. However, our sensitivity analyses revealed that this baseline prognostic risk is largely subsumed by the established Integrative Clusters (INTCLUST) taxonomy. This indicates that while GDBI captures a fundamental aspect of tumour biology, its utility as a standalone baseline prognostic marker is limited in the context of existing comprehensive molecular stratifications.

Despite this overlap in baseline prognostication, GDBI demonstrated distinct potential as a treatment-conditioning variable. Tumours exhibiting the most profound buffering collapse showed a trend toward improved survival following chemotherapy. Although this interaction reached statistical significance in our largest independent accrual sub-cohort, rigorous permutation testing confirmed that the effect does not achieve global significance across the pooled, heterogeneous population.

Ultimately, these findings establish global dosage-buffering capacity as a measurable, epigenetically linked axis of transcriptional resilience. While it does not supersede existing taxonomies for baseline risk assessment, the macroscopic failure of dosage control represents a distinct biological vulnerability. With further validation in targeted, highly unstable tumour sub-populations, quantifying this buffering collapse may provide a valuable tool for refining chemotherapy response prediction in breast cancer.