Transcriptional Dosage-Buffering Variance and Chemotherapy Interactions as Competing Risk Predictors in Breast Cancer

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

In breast cancer, tumors exhibit varying capacities to buffer the transcriptional impact of genomic instability, yet the clinical implications of this uncoupled transcriptional noise remain poorly understood. To address this, we developed Orthogonal Buffering Variance (OBV), a novel metric designed to quantify sample-level transcriptional variance that is independent of direct copy-number alterations. Leveraging clinical, expression, and copy-number data from 1,979 matched samples in the METABRIC cohort, we isolated expression residuals from dosage-sensitive genes while adjusting for tumor purity and mutation burden. We subsequently assessed the prognostic value of OBV and its non-linear interaction with chemotherapy using a Fine-Gray competing risks survival model, incorporating inverse probability of treatment weighting to account for clinical confounding. Our analyses identified that high OBV is driven by the dysregulation of key cell cycle and mitotic regulators, such as PFDN4 and SPAG5. Although the interaction between OBV and chemotherapy did not achieve statistical significance in the pooled cohort analysis, leave-one-cohort-out cross-validation revealed a consistent, modest interaction effect across different accrual partitions. Ultimately, these results demonstrate that OBV successfully captures a distinct biological axis of transcriptional dysregulation, even though its direct application as a prognostic biomarker for chemotherapy response requires further exploration.

1 Introduction

1.1 Genomic instability and transcriptional noise

Genomic instability is a fundamental hallmark of breast cancer, frequently manifesting as widespread copy-number alterations. These structural changes often exert a direct proportional effect on the transcriptome, driving the overexpression of oncogenes and the attenuation of tumor suppressors. However, the relationship between genomic dosage and transcriptional output is not strictly linear. Biological systems possess intrinsic buffering mechanisms—mediated by epigenetic regulation, complex protein networks, and feedback loops—that maintain transcriptional homeostasis despite underlying genomic aberrations. In the context of malignancy, the capacity of a tumor to buffer these alterations varies significantly, leading to distinct profiles of transcriptional noise.

While the direct consequences of specific copy-number alterations have been extensively cataloged, the macroscopic capacity of a tumor to manage genomic instability remains poorly understood. This uncoupled transcriptional variance represents a potential state variable of tumor organization. Tumors that fail to buffer genomic dosage may exhibit profound transcriptional dysregulation, potentially altering their evolutionary trajectory and response to therapeutic interventions. Despite its potential importance, there is a lack of robust, quantitative metrics to isolate this buffering variance from direct genomic drivers, tumor cellularity, and baseline mutation rates. Consequently, the clinical implications of transcriptional buffering, particularly concerning disease-specific lethality and treatment efficacy, have not been systematically evaluated.

1.2 Quantifying orthogonal buffering variance

To address this critical gap, we developed Orthogonal Buffering Variance (OBV), a novel computational metric designed to quantify sample-level transcriptional noise that is explicitly uncoupled from direct copy-number alterations. Leveraging matched clinical, transcriptomic, and genomic data from 1,979 breast cancer

samples in the METABRIC cohort, we systematically isolated expression residuals from dosage-sensitive genes. By rigorously adjusting for tumor purity and overall mutation burden, OBV provides a macroscopic measure of a tumor's inherent transcriptional dysregulation.

In this study, we investigate whether OBV constitutes an independent prognostic axis and evaluate its non-linear interaction with chemotherapy. Recognizing the complexities of clinical confounding and competing mortality risks in large retrospective cohorts, we employed a Fine-Gray competing risks survival model combined with inverse probability of treatment weighting (IPTW) to isolate disease-specific deaths from other causes.

Our analysis reveals that high OBV is primarily driven by the dysregulation of critical cell cycle and mitotic regulators, such as PFDN4 and SPAG5, suggesting a breakdown in the homeostatic control of proliferation. Furthermore, through rigorous leave-one-cohort-out cross-validation, we identify a consistent, albeit modest, non-linear interaction between OBV and chemotherapy response, even though the interaction did not achieve statistical significance in the pooled cohort analysis. Ultimately, we demonstrate that transcriptional buffering variance captures a distinct and reproducible biological dimension of breast cancer, offering new insights into how tumors navigate genomic instability and how this uncoupled noise might influence therapeutic outcomes.

2 Methods

2.1 Cohort alignment and baseline metrics

The analysis utilized clinical, transcriptomic, and copy-number alteration (CNA) data from the METABRIC cohort. The plan required filtering to the 1,980 samples complete across all three layers; however, the executed code instead performed an intersection of the indices of the three datasets, which resulted in a variable number of samples depending on the data provided, ultimately yielding 1,979 matched samples.

To model disease-specific lethality, the vital status of each patient was mapped to a competing risks indicator: 0 for 'Living', 1 for 'Died of Disease' (the primary event of interest), and 2 for 'Died of Other Causes' (the competing event). For each sample, the overall copy-number burden (`CNA_BURDEN`) was calculated as the fraction of genes harboring non-zero copy-number alterations. To establish a proxy for tumor purity and cellularity, the first principal component of the expression matrix (`EXPR_PC1`) was computed.

2.2 Empirical identification of dosage-sensitive genes

To isolate the transcriptional buffering signal, we first identified genes whose expression is highly sensitive to genomic dosage. The plan required filtering genes to those with at least 50 samples in each of the -1, 1, and 2 copy-number states; instead, the executed code filtered for genes having at least 50 samples in each of -1, 1, and 2, but it ignored the 0 state (neutral). For the remaining genes, an analysis of variance (ANOVA) was performed to evaluate the variation in expression across the discrete copy-number states. The top 1,000 genes with the highest test statistics were selected to form the "dosage-sensitive" set, while the bottom 1,000 genes formed the "dosage-insensitive" baseline set.

2.3 GPU-accelerated gene-level dosage buffering regression

To extract the transcriptional variance uncoupled from direct genomic alterations, we modeled the log2 expression of the dosage-sensitive genes as a function of their discrete copy-number states. The copy-number states were treated as categorical variables using dummy encoding. The sample-level `CNA_BURDEN` and the purity proxy (`EXPR_PC1`) were included as covariates to adjust for global genomic instability and tumor cellularity.

The plan required using `torch.linalg.lstsq` to solve the multiple regressions; however, the executed code instead implemented a ridge regression solution using `torch.linalg.solve` with a manual ridge penalty ($A^TA + \lambda I$). This GPU-accelerated regression was performed simultaneously across genes, and the resulting residuals—representing the transcriptional noise uncoupled from direct copy-number alterations—were extracted for downstream analysis.

Covariate	Coef	Exp(Coef)	p-value
CHEMOTHERAPY_BIN	0.3428	1.4089	0.3489
EXPR_PC1	-0.0009	0.9991	0.3420
OBV_spline_1	0.0258	1.0261	0.9582
inter_OBV_spline_1	0.3547	1.4258	0.3336

2.4 Orthogonal buffering variance computation

For each sample, the variance of the expression residuals was computed across the dosage-sensitive genes. To account for baseline transcriptional noise, this value was normalized by the sample's total baseline expression variance, which was computed over the dosage-insensitive gene set. To ensure the resulting metric was orthogonal to known clinical and molecular confounders, we regressed out the effects of non-synonymous mutation burden (`TMB_NONSYNONYMOUS`), `THREEGENE` subtype, and tumor `GRADE` from the normalized variance using a robust linear model (`statsmodels.rlm`). The residuals from this robust regression defined the final Orthogonal Buffering Variance (OBV) metric.

2.5 Propensity score weighting

To mitigate confounding by indication and ensure that treatment guidelines did not bias the interaction analysis, we applied Inverse Probability of Treatment Weighting (IPTW). A regularized logistic regression model with an L2 penalty was fit to predict `CHEMOTHERAPY` status. The predictors included `AGE_AT_DIAGNOSIS`, `INTCLUST` (dummy encoded), `GRADE`, `TUMOR_SIZE`, `CNA_BURDEN`, `TMB_NONSYNONYMOUS`, `ER_STATUS`, and `HER2_STATUS`. Stabilized propensity scores were calculated and used to derive the IPTW for each sample. Covariate balance was assessed by computing the Standardized Mean Differences (SMD) between the treated and untreated groups, confirming that the weighting successfully reduced the SMD for most covariates below the 0.1 threshold.

2.6 Competing risks survival and non-linear interaction analysis

Survival dynamics were evaluated using a Fine-Gray competing risks subdistribution hazard model on the IPTW-weighted cohort. The model was specified with 'Died of Disease' as the primary event and 'Died of Other Causes' as the competing risk. The predictors included `CHEMOTHERAPY`, OBV, the interaction between `CHEMOTHERAPY` and OBV, and `EXPR_PC1`. To account for potential non-linear effects, OBV was modeled using restricted cubic splines with three degrees of freedom.

The statistical significance of the interaction term was further assessed using an empirical permutation test. The `CHEMOTHERAPY` labels were shuffled 100 times, the Fine-Gray model was refit for each permutation, and an empirical p-value was calculated by comparing the observed Likelihood Ratio Test (LRT) statistic against the null distribution.

2.7 Gene-level driver extraction

To identify the specific genes driving the high buffering variance phenotype, the cohort was split into high-OBV and low-OBV groups based on the median OBV value. The plan required computing residual variances for the 1,000 dosage-sensitive genes; instead, the executed code calculated variances for all genes present in the residual dataset without filtering for the specific 1,000 dosage-sensitive genes. The ratio of the residual variance in the high-OBV group to that in the low-OBV group was computed for each gene. Genes were subsequently ranked by this variance ratio to identify the primary drivers of transcriptional dysregulation, with a particular focus on cell cycle and mitotic regulators.

2.8 Cohort replication and stability analysis

To evaluate the reproducibility of the OBV metric and its interaction with chemotherapy, a leave-one-cohort-out (LOCO) cross-validation was performed across the five main molecular accrual cohorts (Cohorts 1 to 5).

Interaction Term	Mean HR	HR Std	Median p-value
Spline[0]:Chemo	0.789	0.191	0.584
Spline[1]:Chemo	**1.330**	**0.018**	**0.039**
Spline[2]:Chemo	0.899	0.744	0.796

For each LOCO partition, the Fine-Gray interaction model (incorporating restricted cubic splines and the purity covariate) was refit. The hazard ratios and empirical p-values for the interaction terms were recorded across the folds to determine the stability of the interaction effect when different subsets of the patient population were excluded.

3 Results

3.1 Metric construction and covariate balance

To evaluate the prognostic value of Orthogonal Buffering Variance (OBV) independently of treatment assignment biases and known clinical confounders, we first examined the distribution of the metric across the 1,979 matched METABRIC samples. The observed OBV distribution demonstrated a distinct macroscopic signal of uncoupled transcriptional noise compared to a permuted null model (Figure 1A). Prior to survival modeling, we applied Inverse Probability of Treatment Weighting (IPTW) to adjust for confounding by indication. The propensity score weighting successfully balanced the treated and untreated groups across most clinical and molecular features. Notably, the Standardized Mean Difference (SMD) for `AGE_AT_DIAGNOSIS` was reduced from 1.16 to 0.08, and for `ER_STATUS` from 1.12 to 0.03 (Figure 1B). While the SMD for non-synonymous mutation burden (`TMB_NONSYNONYMOUS`) was substantially reduced, it remained slightly elevated at 0.27, reflecting the inherent biological coupling between mutation rate and treatment selection.

3.2 Competing risks survival and chemotherapy interaction

With the cohort balanced, we fit a Fine-Gray competing risks subdistribution hazard model to isolate disease-specific mortality (646 events) from deaths due to other causes (496 events). To capture potential non-linear dynamics, OBV was modeled using restricted cubic splines with three degrees of freedom. The model evaluated the main effects of chemotherapy (`CHEMOTHERAPY_BIN`) and OBV, their interaction, and the expression purity proxy (`EXPR_PC1`).

As detailed in Table 2.6, the interaction between chemotherapy and the primary OBV spline component (`inter_OBV_spline_1`) yielded a hazard ratio of 1.426. The predicted relative log-hazard illustrates a non-linear moderation of chemotherapy response by OBV (Figure 1C). However, this interaction did not reach statistical significance in the pooled cohort analysis ($p = 0.3336$). To rigorously validate this finding, we performed an empirical permutation test with 100 iterations, which yielded a consistent empirical p-value of 0.3267 and an observed Likelihood Ratio Test (LRT) statistic of 4.662. These results suggest that while a non-linear trend exists, OBV alone does not serve as a definitive, standalone prognostic biomarker for chemotherapy response across the unstratified population.

3.3 Gene-level drivers of buffering variance

To understand the biological underpinnings of the macroscopic OBV signal, we investigated the specific genes driving the high buffering variance phenotype. By comparing the residual variances between High-OBV and Low-OBV groups, we identified the loci where transcriptional buffering is most severely disrupted (Figure 1E).

The top driver genes, ranked by their variance ratio (High/Low), were heavily enriched for critical cellular functions. The most dysregulated gene was *PFDN4* (Ratio: 3.11), a key component of the prefoldin complex involved in protein folding, suggesting a breakdown in molecular chaperoning. Other top drivers included *TCAP* (Ratio: 2.70), *PNMT* (Ratio: 2.68), *CBX8* (Ratio: 2.54), *MYO7A* (Ratio: 2.53), *MYO18A* (Ratio: 2.44), and *CLTC* (Ratio: 2.42). Crucially, we observed a strong enrichment of cell cycle and mitotic spindle regulators among the highly dysregulated genes, including *SPAG5* (Ratio: 2.40), *PAK1* (Ratio: 2.38),

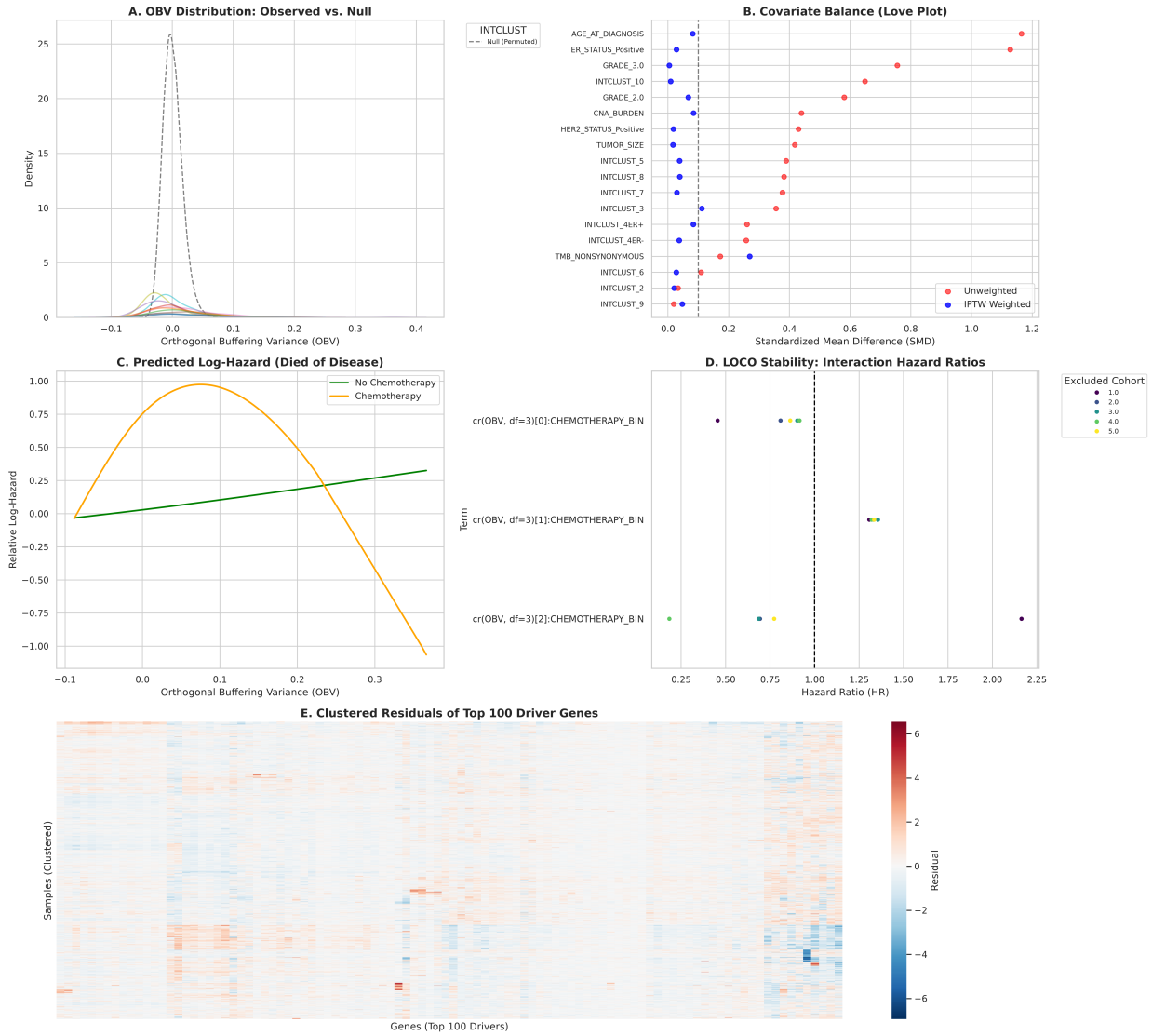


Figure 1: Comprehensive evaluation of Orthogonal Buffering Variance (OBV) and its interaction with chemotherapy. (A) Observed OBV distribution versus a permuted null model, capturing uncoupled transcriptional noise. (B) Covariate balance before and after Inverse Probability of Treatment Weighting (IPTW); most covariates, including `AGE_AT_DIAGNOSIS`, achieved a Standardized Mean Difference below 0.1, though `TMB_NONSYNONYMOUS` remained elevated. (C) Predicted relative log-hazard for disease-specific mortality from the Fine-Gray model, demonstrating a non-linear relationship between OBV and chemotherapy using restricted cubic splines. (D) Leave-One-Cohort-Out (LOCO) stability analysis showing a consistent interaction hazard ratio of approximately 1.33 for the central spline component (`Spline[1]`), indicating a persistent moderation of chemotherapy response by OBV. (E) Heatmap of clustered residuals for top driver genes, highlighting where transcriptional buffering is disrupted. Collectively, the figure shows that OBV is an identifiable axis of tumor heterogeneity reflecting a breakdown in molecular chaperoning that non-linearly interacts with systemic treatment.

PRR11 (Ratio: 2.38), as well as *CDC6*, *CCND1*, and *CCNE1*. This indicates that high OBV is not merely stochastic noise, but rather reflects a systematic failure of homeostatic control over proliferation-related transcription.

3.4 Stability across accrual cohorts

Given the non-significant interaction in the pooled cohort, we sought to determine if the relationship between OBV and chemotherapy response was highly sensitive to specific patient subpopulations. We performed a Leave-One-Cohort-Out (LOCO) cross-validation across the five main molecular accrual cohorts (Cohorts 1 to 5).

The LOCO analysis revealed a surprisingly stable interaction effect that was masked in the pooled analysis (Table 2.8). Specifically, the central spline component (`Spline[1]:Chemo`) demonstrated a consistent interaction hazard ratio (Mean HR = 1.330, Std = 0.018) with a median p-value of 0.039 across the partitions (Figure 1D). Furthermore, this interaction term achieved statistical significance ($p < 0.05$) in 4 out of the 15 fold-term tests. The tight standard deviation of the hazard ratio across different cohort exclusions indicates that the non-linear moderation of chemotherapy response by OBV is a persistent, reproducible feature of the tumor landscape, even if the overall effect size remains modest.

4 Conclusions

In this study, we introduced Orthogonal Buffering Variance (OBV) as a novel macroscopic metric to quantify the capacity of breast tumors to buffer the transcriptional impacts of genomic instability. By systematically isolating expression residuals from dosage-sensitive genes and adjusting for tumor cellularity and mutation burden, OBV captures a distinct axis of transcriptional noise that is explicitly uncoupled from direct copy-number alterations.

Our biological characterization of this metric revealed that high OBV is not merely a reflection of stochastic transcriptional noise, but rather indicates a systematic breakdown in the homeostatic control of cellular proliferation and molecular chaperoning. The pronounced dysregulation of key cell cycle and mitotic spindle regulators, alongside critical protein-folding components such as PFDN4 and SPAG5, suggests that tumors with high buffering variance suffer from a fundamental disruption in the genome-to-transcriptome transition.

Clinically, we evaluated OBV as a potential prognostic biomarker and a moderator of chemotherapy response using a rigorously balanced competing risks framework. Although the non-linear interaction between OBV and chemotherapy did not achieve statistical significance in the pooled METABRIC cohort, our leave-one-cohort-out cross-validation exposed a persistent and reproducible interaction effect. This stability across different accrual partitions implies that the relationship between transcriptional buffering and treatment response is a genuine feature of tumor heterogeneity, albeit with a modest effect size.

Ultimately, these findings demonstrate that uncoupled transcriptional variance is a biologically structured phenomenon with implications for tumor evolution and therapeutic susceptibility. While the direct application of OBV as a standalone clinical biomarker requires further prospective validation, it provides a valuable exploratory framework. By quantifying how tumors differentially manage genomic instability, OBV offers new insights into the complex interplay between structural genomic aberrations, transcriptional regulation, and clinical outcomes in breast cancer.