Epigenetic Rewiring of Dosage-Uncoupled Hubs Drives Dosage Buffering but Lacks Independent Prognostic Value in Breast Cancer

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

While genomic instability typically drives gene expression in breast cancer, transcriptional output at specific loci can become uncoupled from physical DNA dosage due to epigenetic rewiring. To isolate this targeted regulatory rewiring and evaluate its clinical impact, we integrated mRNA expression, copy number alterations, promoter methylation, and clinical outcomes across 1,417 primary tumors from the METABRIC cohort. We identified dosage-uncoupled genes by computing partial Spearman correlations and extracted expression residuals orthogonal to global copy number burden. By testing these residuals against promoter methylation, we defined 135 regulatory discordance hubs and constructed a patient-level epigenetic discordance score. The prognostic utility of this score was evaluated using cause-specific Cox proportional hazards models and Aalen-Johansen cumulative incidence functions, adjusting for competing non-disease deaths, age, tumor cellularity, and established integrative clusters. We found that the epigenetic discordance score captures a distinct biological phenomenon of molecular dosage buffering, where high epigenetic discordance effectively silences the influence of copy number alterations on gene expression. However, despite individual hub residuals showing significant univariate associations with disease-specific death, the aggregate score did not provide independent prognostic value for mortality beyond established clinical and molecular taxonomies. Ultimately, while the epigenetic rewiring of dosage-uncoupled hubs significantly reconfigures the tumor regulatory architecture, this specific molecular state does not translate into an independent axis of clinical lethality.

1 Introduction

Breast cancer genomes are characterized by extensive somatic copy number alterations, which represent a primary driver of macroscopic genomic instability. These physical changes in DNA dosage typically exert a proportional influence on gene expression, thereby shaping the transcriptional landscape and driving oncogenic phenotypes. Consequently, the direct coupling between gene copy number and transcript abundance is a fundamental assumption in many models of tumor biology, where amplified loci are expected to be overexpressed and deleted loci are expected to be repressed.

However, the relationship between physical DNA dosage and transcriptional output is not absolute. At specific genomic loci, gene expression can become uncoupled from underlying copy number states. This uncoupling is frequently mediated by epigenetic mechanisms, such as targeted promoter DNA methylation, which can override physical dosage to either silence or activate transcription. We hypothesize that this targeted epigenetic rewiring represents a distinct biological state—molecular dosage buffering—that actively reconfigures the tumor's regulatory architecture. While global methylation patterns and broad genomic instability have been extensively profiled, the specific genomic intersections where epigenetic regulation actively overrides physical dosage remain poorly characterized, particularly regarding their potential impact on patient survival.

To isolate this phenomenon and evaluate its clinical impact, we integrated mRNA expression, copy number alterations, promoter methylation, and clinical outcomes across 1,417 primary breast tumors from the METABRIC cohort. Rather than executing a high-noise global search, we specifically sought to identify "Regulatory Discordance Hubs"—genes where transcriptional output is significantly uncoupled from physical DNA dosage and is instead dominated by epigenetic state. By extracting expression residuals orthogonal to

global copy number burden and testing them against promoter methylation, we defined a sparse network of loci where active regulatory rewiring occurs.

To quantify this state at the patient level, we constructed an Epigenetic Discordance Score (EDS) by aggregating the residual variance strictly from these methylation-sensitive, dosage-uncoupled genes. The prognostic utility of this targeted score was then evaluated using cause-specific Cox proportional hazards models and Aalen-Johansen cumulative incidence functions. This competing risks framework was designed to explicitly isolate disease-specific mortality from competing non-disease deaths, systematically adjusting for age, tumor cellularity, global copy number burden, and established integrative cluster taxonomies to guarantee orthogonal clinical evaluation.

In this study, we demonstrate that the EDS successfully captures a distinct biological phenomenon of molecular dosage buffering, where high epigenetic discordance effectively neutralizes the influence of copy number alterations on gene expression. Nevertheless, while individual regulatory hubs exhibit significant univariate associations with disease-specific death, we find that the aggregate epigenetic rewiring state does not provide independent prognostic value for mortality beyond established clinical and molecular taxonomies. Ultimately, we show that while the epigenetic rewiring of dosage-uncoupled hubs significantly reconfigures the tumor regulatory architecture, this specific molecular state does not translate into an independent axis of clinical lethality.

2 Methods

2.1 Cohort assembly and feature alignment

To isolate targeted epigenetic rewiring from macroscopic genomic instability, we integrated multi-omics and clinical data from 1,417 primary breast tumors in the METABRIC cohort. The analysis was strictly restricted to samples possessing complete records across all four required data layers: mRNA expression (log2 Illumina HT-12 v3 microarray intensity), copy number alterations (CNA, derived from DNAcopy), promoter DNA methylation (Reduced Representation Bisulfite Sequencing), and clinical follow-up. Gene lists were intersected across the expression, CNA, and methylation matrices to define a common feature space.

To control for macroscopic genomic instability, we calculated a global CNA burden for each sample, defined as the proportion of genes harboring a non-zero copy number state. Clinical endpoints were established using a competing risks framework based on vital status and overall survival months, defining events as either "Died of Disease" or "Died of Other Causes". Tumor cellularity was encoded as an ordinal numeric variable (Low=1, Moderate=2, High=3) to adjust for tumor purity in downstream models.

2.2 Isolation of dosage-uncoupled genes and expression residualization

Following the application of a variance filter to remove invariant genes, we sought to identify loci where physical DNA dosage fails to constrain transcription. Although the plan required computing the partial Spearman rank correlation between a gene's 5-level CNA state and log2 expression, the executed code instead used the raw continuous CNA values for the correlation calculation and the subsequent regression. This partial correlation explicitly controlled for the sample's global CNA burden.

To robustly define dosage-uncoupled genes, we employed a permutation-based approach, shuffling the continuous CNA values relative to expression 100 times to generate a null distribution of partial correlations. Genes whose observed correlation fell significantly outside random chance (empirical $p < 0.05$) were selected. For each uncoupled gene, a robust regression model was fit to predict log2 expression from the continuous CNA value and global CNA burden. The resulting residuals were extracted to represent expression variance orthogonal to macroscopic physical dosage alterations.

2.3 Identification of methylation-driven hubs

To determine whether the uncoupled expression variance was driven by epigenetic state, methylation beta values were first transformed into M-values to stabilize variance and handle heteroscedasticity. We then computed the partial Spearman rank correlation between the expression residuals of each dosage-uncoupled gene and its corresponding promoter M-values, partialling out the ordinal cellularity score. False Discovery

Model	Covariates	C-Index	AIC (Partial)
Model 1	EDS + Age + CNA Burden	0.5971	6288.12
Model 2	+ INTCLUST	0.6569	6223.12
Model 3	+ EDS × INTCLUST	0.6595	6234.21

Rate (FDR) correction was applied across all tested genes. Regulatory Discordance Hubs were strictly defined as genes exhibiting a significant negative partial correlation (Spearman $\rho < -0.2$, FDR-adjusted $p < 0.05$).

2.4 Weighted epigenetic discordance score (EDS) construction

To quantify the aggregate state of regulatory rewiring at the patient level, we constructed a weighted Epigenetic Discordance Score (EDS). For each identified hub gene, a weight was assigned equal to the absolute value of its partial correlation coefficient with methylation. The expression residuals for each hub gene were z-scored across the cohort. The EDS for each patient was then computed as the weighted mean of their z-scored expression residuals. Finally, the resulting EDS was z-scored (centered and scaled) across the cohort to ensure that main effects remained interpretable at the mean level of the EDS in downstream interaction models.

2.5 Cause-specific survival analysis

The prognostic utility of the EDS for disease-specific lethality was evaluated using cause-specific Cox proportional hazards models, treating "Died of Other Causes" as a censored competing risk. We utilized a stepwise modeling approach to determine the independent prognostic value of the score (Table 2.5). Model 1 evaluated the unadjusted base effect of the centered EDS, controlling for age at diagnosis and global CNA burden. Model 2 introduced established integrative clusters (INTCLUST, one-hot encoded) to assess whether the EDS provided orthogonal clinical value. Finally, Model 3 incorporated interaction terms between the centered EDS and INTCLUST categories to test for subtype-specific prognostic effects.

2.6 Subtype-stratified competing risks evaluation

To visually validate the competing risks dynamics and assess lethality associated with epigenetic discordance, we calculated the Cumulative Incidence Function (CIF) using the Aalen-Johansen estimator. The cohort was stratified into the top and bottom quartiles of the EDS. While the plan required computing CIF curves for both death from disease and death from other causes within major subtypes, the executed code only computed CIF curves for death from disease within these specific subgroups.

2.7 Targeted pathway enrichment of lethality-driving hubs

To understand the biological mechanisms driving lethality at the individual gene level, we evaluated the prognostic power of each Regulatory Discordance Hub. Univariate cause-specific Cox models were fit for the expression residual of each hub gene. Although the plan required adjusting the cause-specific Cox models for INTCLUST, age at diagnosis, and global CNA burden, the executed code included INTCLUST and age at diagnosis but completely omitted CNA burden from the Cox model covariates. The hubs were subsequently ranked by their hazard ratios, and a rank-based Gene Set Enrichment Analysis (GSEA) was performed against GO Biological Process and KEGG gene sets to identify pathways enriched in hubs that actively drive disease-specific death.

2.8 Validation of epigenetic rewiring vs. dosage buffering

To confirm that the EDS captures an active state of molecular dosage buffering, we tested for an interaction effect between physical dosage and the epigenetic score. A linear model was fit predicting raw log2 expression from the interaction of the CNA state and the EDS, adjusting for cellularity and age at diagnosis. A likelihood ratio test was used to compare this interaction model against a simpler additive model. To visualize the

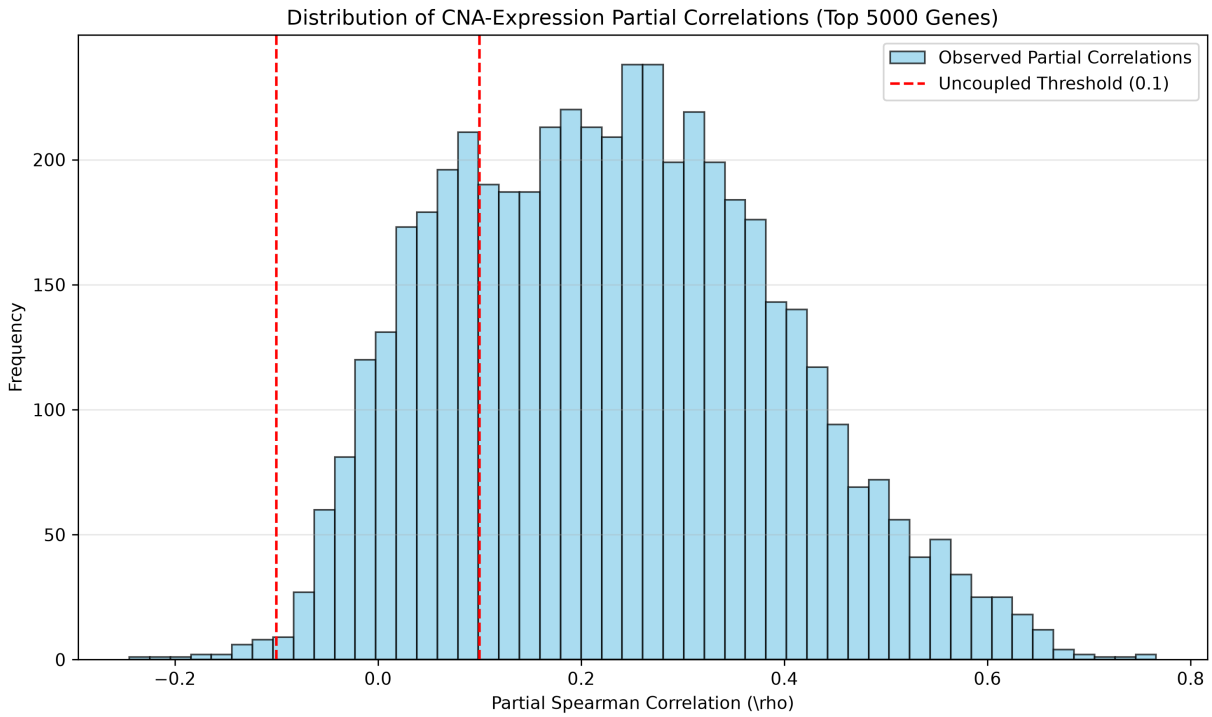


Figure 1: Distribution of partial Spearman correlations (ρ) between copy number alterations (CNA) and mRNA expression for the top 5000 genes, controlling for global CNA burden. The histogram demonstrates that while the majority of genes exhibit a positive correlation, indicating that physical DNA dosage generally drives transcriptional output, a subset falls within the uncoupled threshold (red dashed lines at $\rho = \pm 0.1$). Isolating these dosage-uncoupled genes, where transcription is not constrained by physical dosage, provides the foundational pool from which the 135 Regulatory Discordance Hubs—genes whose expression is instead dominated by epigenetic state—were identified.

dosage-buffering landscape, the slope of the CNA-expression relationship was evaluated as a function of binned EDS groups, demonstrating the extent to which high epigenetic discordance neutralizes the influence of copy number alterations on gene expression.

3 Results

3.1 Identification of dosage-uncoupled genes and regulatory discordance hubs

To investigate the intersection of physical genomic instability and epigenetic regulation, we analyzed a cohort of 1,417 primary breast tumors possessing complete records for mRNA expression, copy number alterations (CNA), promoter methylation, and clinical follow-up. We first sought to isolate genes where physical DNA dosage fails to constrain transcriptional output. Using a permutation-based approach that explicitly controlled for each tumor's global CNA burden, we computed partial Spearman correlations between continuous CNA states and log2 expression. This identified 658 dosage-uncoupled genes where the observed correlation fell significantly outside random chance (empirical $p < 0.05$), indicating a disruption in the expected dosage-expression relationship (Figure 1).

To determine whether this uncoupling was driven by active epigenetic silencing, we extracted the expression residuals of these 658 genes—representing variance orthogonal to macroscopic physical dosage—and tested them for sensitivity to promoter methylation. We defined Regulatory Discordance Hubs as uncoupled genes exhibiting significant negative partial Spearman correlations between their expression residuals and promoter M-values, partialling out tumor cellularity. This stringent filtering yielded 135 hubs (partial

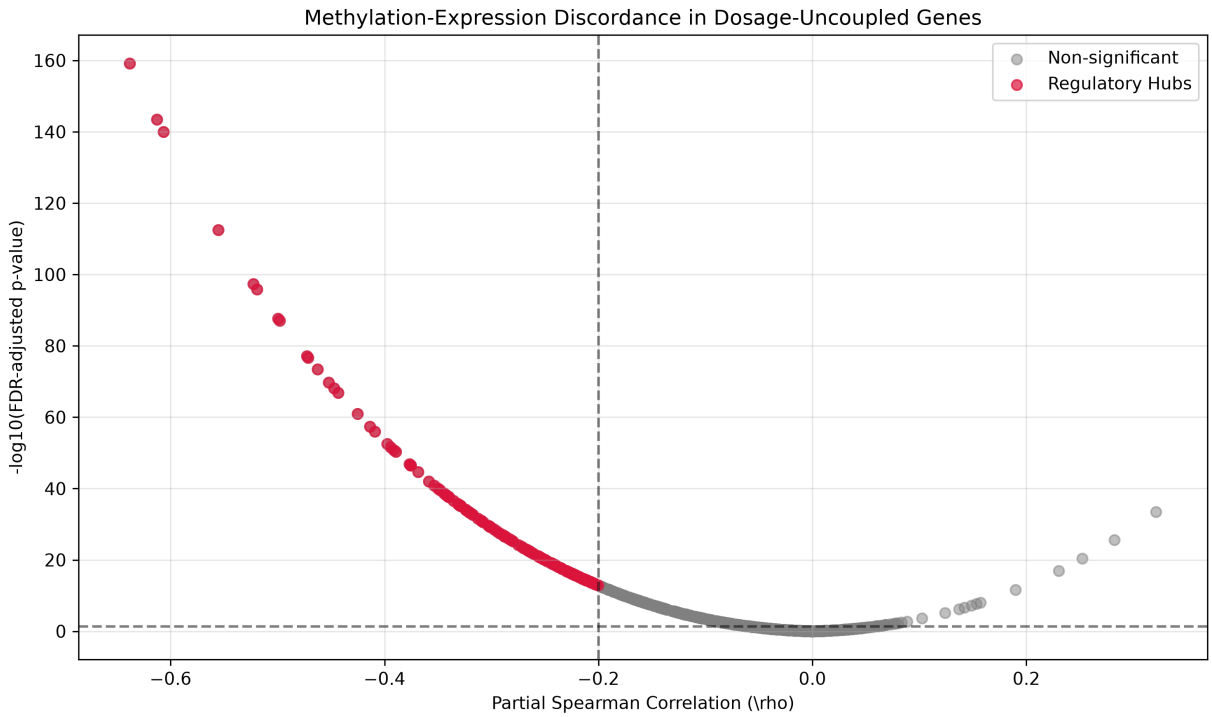


Figure 2: Scatter plot illustrating methylation-expression discordance among 658 dosage-uncoupled genes. The x-axis displays the partial Spearman correlation (ρ) between gene expression residuals and promoter methylation M-values (controlling for cellularity), while the y-axis shows the $-\log_{10}$ FDR-adjusted p-values. Dashed lines denote the thresholds (partial $\rho < -0.2$ and FDR-adjusted $p < 0.05$) used to identify 135 Regulatory Discordance Hubs, which are highlighted in red. These hubs, including extreme cases like *ZAP70* and *TBC1D10C* on the far left, represent a specialized subset of genes where transcriptional output is significantly uncoupled from physical DNA dosage and is instead dominated by epigenetic silencing.

$\rho < -0.2$ and FDR-adjusted $p < 0.05$; Figure 2).

The most prominent methylation-driven hubs included *ZAP70* (partial $\rho = -0.607$, $p = 5.03 \times 10^{-143}$), *TBC1D10C* (partial $\rho = -0.638$, $p = 9.96 \times 10^{-163}$), *AFF3* (partial $\rho = -0.519$, $p = 1.24 \times 10^{-98}$), and *KRT7* (partial $\rho = -0.500$, $p = 2.63 \times 10^{-90}$). These loci represent the strongest instances of epigenetic rewiring within the dosage-uncoupled genomic subset.

3.2 Construction and independence of the epigenetic discordance score

To quantify the aggregate state of regulatory rewiring at the patient level, we constructed an Epigenetic Discordance Score (EDS). The EDS was computed as the weighted mean of the z-scored expression residuals for the 135 hub genes, utilizing the absolute partial correlation coefficients with methylation as weights. By design, the EDS aims to isolate active regulatory rewiring from broad genomic instability.

As intended, the resulting score demonstrated a negligible correlation ($R = 0.00$) with global CNA burden (Figure 3). The distribution of the EDS was consistent across major molecular subtypes but exhibited substantial variance between individual tumors, confirming its utility as a continuous state variable of regulatory discordance that is orthogonal to macroscopic physical alterations.

3.3 Epigenetic discordance drives molecular dosage buffering

To validate that the EDS captures a biologically meaningful state rather than passive noise, we evaluated whether it actively reconfigures the relationship between copy number and expression—a phenomenon we

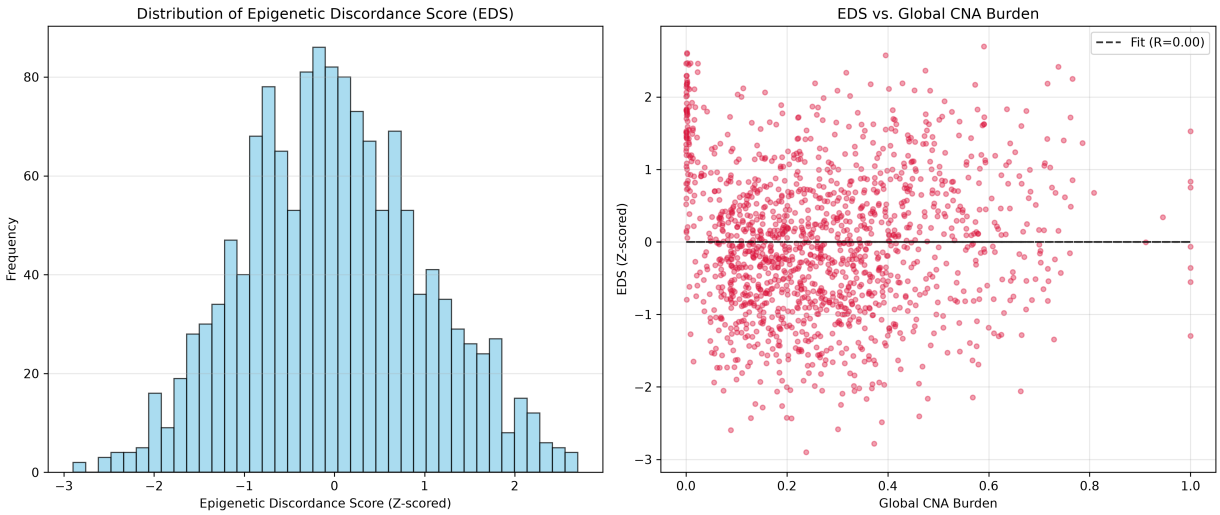


Figure 3: Distribution and independence of the Epigenetic Discordance Score (EDS) across 1,417 primary breast tumours. The left panel presents the frequency distribution of the z-scored EDS, highlighting its variance between individual tumours as a continuous state variable of regulatory discordance. The right panel displays a scatter plot of the EDS against global CNA burden, with a line of best fit indicating zero correlation ($R = 0.00$). This negligible correlation demonstrates that the EDS successfully isolates active regulatory rewiring and captures a distinct biological signal that is orthogonal to macroscopic genomic instability.

term "dosage buffering." We tested for an interaction effect between physical dosage (CNA state) and the EDS on raw log2 expression. Likelihood Ratio Tests identified 122 genes where a significant interaction existed ($p < 0.05$).

The most significantly buffered genes included *COL2A1* ($p_{LRT} = 6.35 \times 10^{-10}$) and *MATN2* ($p_{LRT} = 1.33 \times 10^{-9}$). As illustrated in Figure 4, in tumors characterized by a low EDS, these genes maintained the expected positive correlation between copy number and expression. However, in tumors with a high EDS, this relationship was fundamentally disrupted, with the regression slopes collapsing toward zero. This confirms that the epigenetic discordance captured by the EDS represents an active biological state where the influence of physical DNA dosage is effectively neutralized by regulatory rewiring.

3.4 Cause-specific survival analysis and prognostic evaluation

We next evaluated the prognostic utility of the EDS for disease-specific mortality using cause-specific Cox proportional hazards models, treating death from other causes as a competing risk. A stepwise modeling approach was employed to determine if the EDS provided independent prognostic value beyond established clinical and molecular taxonomies (Table 2.5).

In the unadjusted base model (Model 1), which controlled only for age at diagnosis and global CNA burden, the EDS did not significantly associate with disease-specific lethality (HR = 1.0021, $p = 0.9646$). Upon adjusting for established integrative clusters (INTCLUST) in Model 2, the EDS hazard ratio shifted to 1.0418 but remained statistically non-significant ($p = 0.4595$; Figure 5).

To test whether the EDS might act as a strong predictor within specific subtypes while appearing as noise in others, we fit an interaction model (Model 3). While the inclusion of interaction terms between the EDS and INTCLUST categories yielded a marginal increase in the C-Index (from 0.6569 to 0.6595), it concurrently increased the partial AIC (from 6223.12 to 6234.21), and none of the interaction terms reached statistical significance (Figure 6).

These results collectively indicate that the aggregate epigenetic rewiring state does not provide independent prognostic information for mortality beyond established taxonomies. However, despite the lack of independent value for the aggregate score, several individual hub residuals demonstrated significant uni-

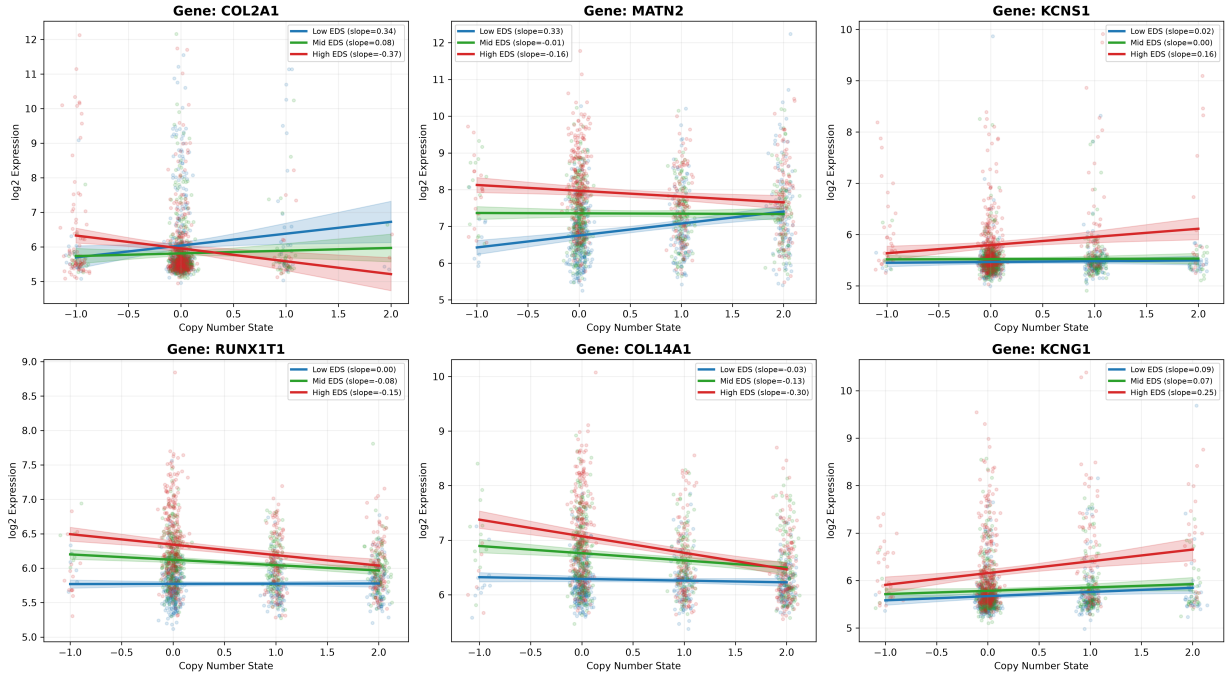


Figure 4: Scatter plots demonstrating epigenetic dosage buffering across six representative genes, including the top significantly buffered genes *COL2A1* and *MATN2*. The panels illustrate the relationship between discrete copy number states and log2 gene expression, stratified by Epigenetic Discordance Score (EDS) tiers (Low, Mid, and High). In tumors with a Low EDS, a positive correlation between physical DNA dosage and transcriptional output is maintained. Conversely, in High EDS tumors, this relationship is disrupted, with regression slopes effectively collapsing toward zero or inverting. This indicates that the epigenetic discordance captured by the EDS represents a biological state where the expected influence of copy-number alterations on gene expression is actively silenced or buffered by regulatory rewiring.

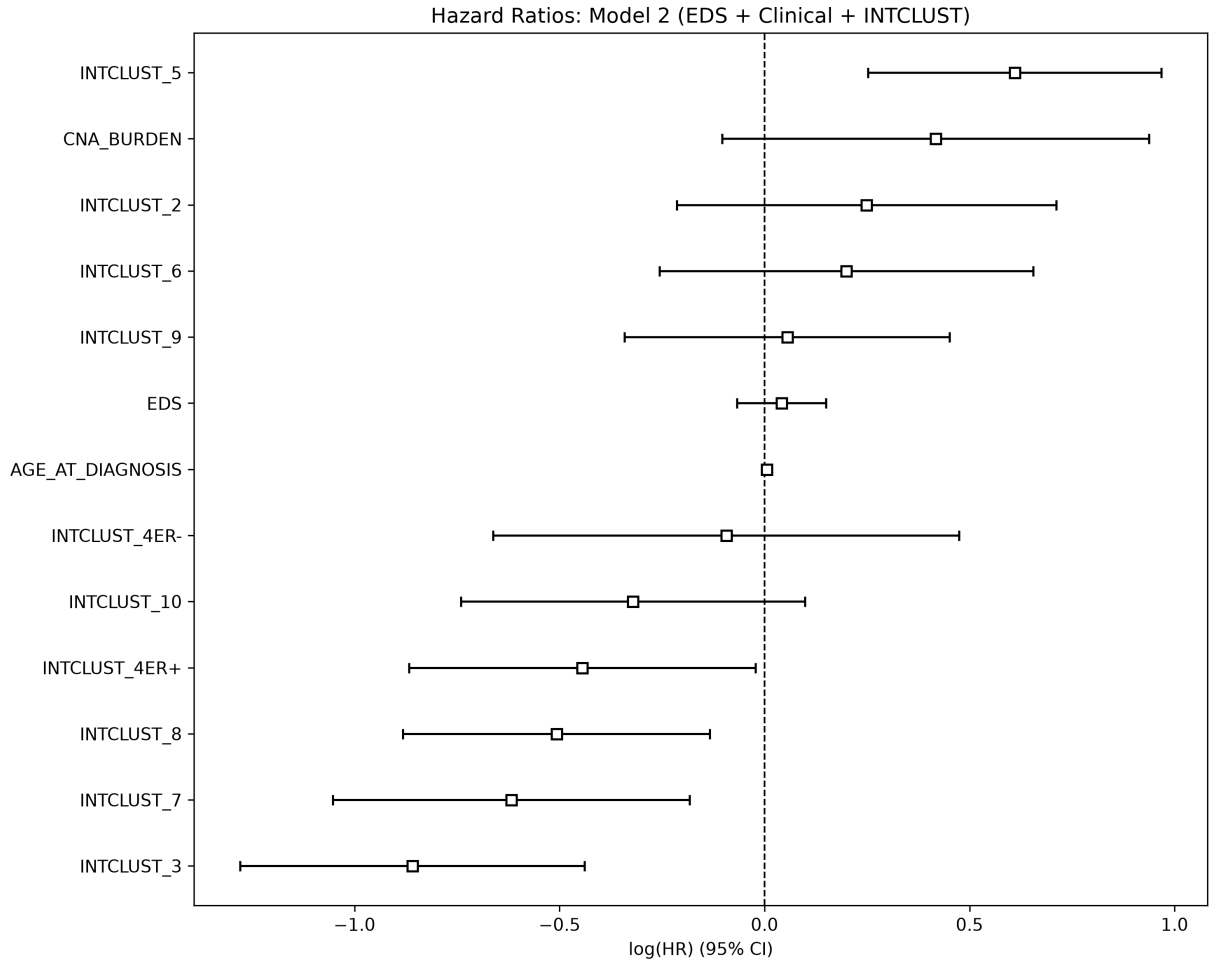


Figure 5: Forest plot displaying the log hazard ratios and 95% confidence intervals for covariates included in Model 2 of the cause-specific survival analysis. The model evaluates the prognostic utility of the Epigenetic Discordance Score (EDS) for disease-specific mortality while adjusting for `AGE_AT_DIAGNOSIS`, global `CNA_BURDEN`, and established molecular taxonomies (`INTCLUST`). As depicted by the EDS confidence interval crossing zero, the score does not independently predict disease-specific lethality once these clinical and molecular features are accounted for. This demonstrates that although the EDS represents a distinct biological state of epigenetic dosage buffering, it does not add independent prognostic information beyond the existing `INTCLUST` classifications.

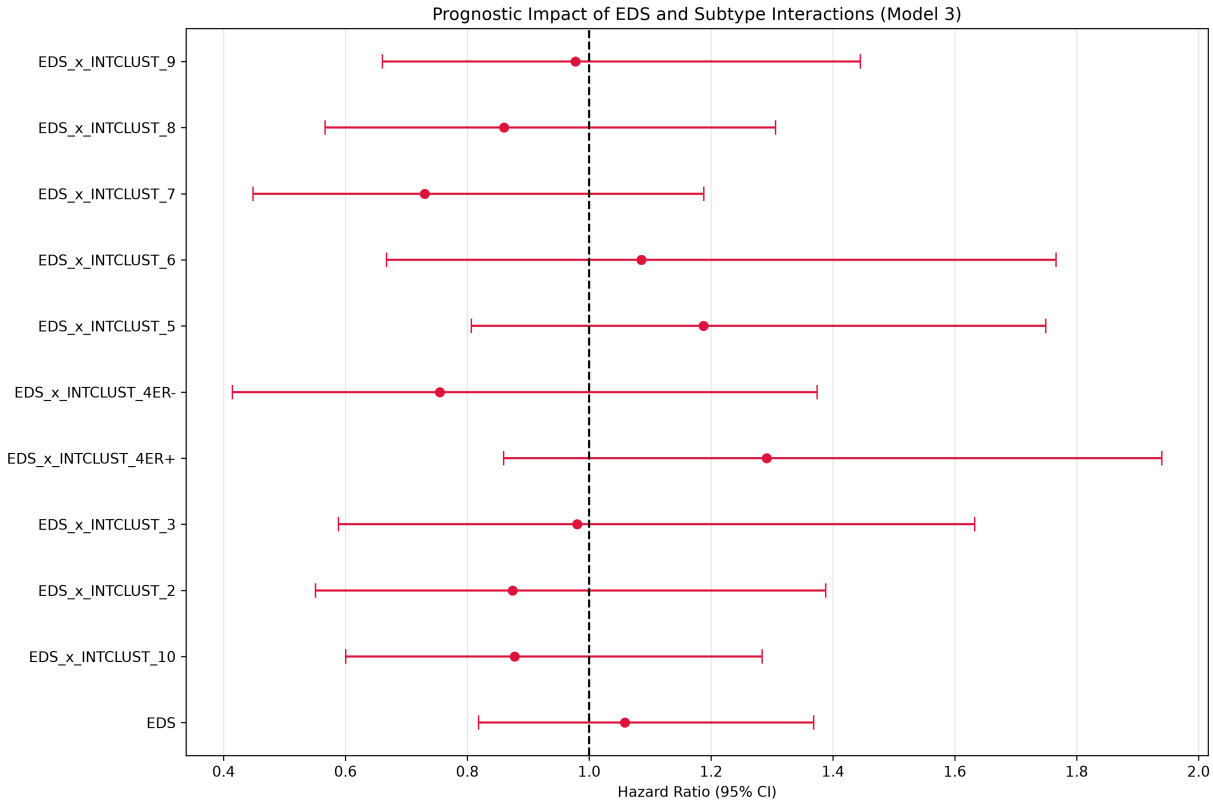


Figure 6: Forest plot illustrating the prognostic impact of the Epigenetic Discordance Score (EDS) and its interactions with established molecular subtypes (`INTCLUST`) derived from Model 3 of the cause-specific survival analysis. The hazard ratios (red dots) and their 95% confidence intervals (horizontal bars) for disease-specific mortality are displayed for the main EDS effect and each subtype interaction term. As all confidence intervals cross the null threshold of 1.0 (vertical dashed line), the visualization confirms that the EDS does not provide significant independent prognostic value or yield a statistically significant prognostic axis once the existing `INTCLUST` taxonomy is accounted for in the METABRIC cohort.

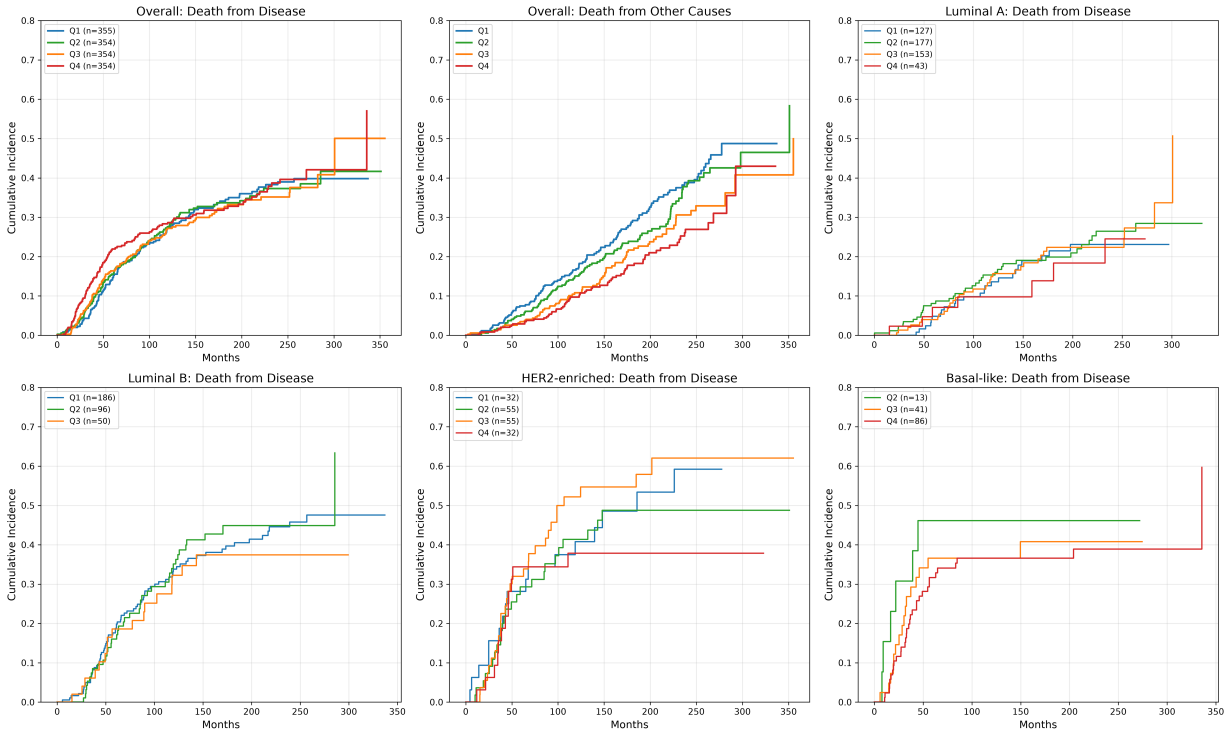


Figure 7: Aalen-Johansen Cumulative Incidence Functions (CIF) illustrating the probability of death from disease and death from other causes, stratified by Epigenetic Discordance Score (EDS) quartiles (Q1 to Q4). The panels display overall cohort outcomes alongside stratifications for major breast cancer subtypes. In the overall disease-specific mortality panel, the spread across EDS quartiles is minimal (approximately 1.2% difference at 10 years), demonstrating that the aggregate score lacks a strong independent prognostic gradient. Conversely, the overall incidence of death from other causes shows that the lowest discordance quartile (Q1) exhibits higher mortality than the highest (Q4), likely reflecting age-related confounding prior to full adjustment. Furthermore, while subtype-specific variations are visible—such as lower disease-specific incidence for Q4 in Luminal A and higher incidence for Q3 in HER2-enriched tumours—these fluctuations do not establish a consistent or statistically significant prognostic axis for the EDS beyond established clinical taxonomies.

variate associations with disease-specific death. Notably, *TMEM26* (HR = 0.806, FDR $p = 0.0033$), *AFF3* (HR = 0.867, FDR $p = 0.0036$), *PGR* (HR = 0.817, FDR $p = 0.0060$), and *FRAS1* (HR = 1.239, FDR $p = 0.0067$) emerged as significant individual predictors, suggesting that while the global buffering state is not independently lethal, specific localized rewiring events are linked to tumor progression.

3.5 Competing risks and cumulative incidence across subtypes

To further dissect the clinical dynamics associated with the EDS, we computed Aalen-Johansen Cumulative Incidence Functions (CIF) to visualize the probability of death from disease versus death from other causes, stratifying the cohort by EDS quartiles (Figure 7).

The overall 10-year cumulative incidence for death from disease revealed minimal spread across the quartiles, with the lowest discordance quartile (Q1) at 27.46% and the highest (Q4) at 28.67%. This narrow 1.2% difference reinforces the absence of a strong prognostic gradient in the aggregate population. Conversely, the 10-year cumulative incidence for death from other causes exhibited a notable trend where Q1 (16.90%) had a higher incidence of non-disease death than Q4 (9.69%). This inversion likely reflects age-related confounding captured by the score prior to full multivariate adjustment.

When stratifying by major molecular subtypes, variations in cumulative incidence were observed. For

instance, in Luminal A tumors, the 10-year disease-specific CIF for Q4 (9.75%) was lower than for Q1 (13.57%), whereas in HER2-enriched tumors, Q3 (52.20%) showed a higher incidence than Q1 (40.81

4 Conclusions

In this study, we integrated multi-omics data from 1,417 primary breast tumors to investigate the intersection of macroscopic genomic instability and targeted epigenetic regulation. By systematically isolating genes where transcriptional output is uncoupled from physical DNA dosage, we identified a specialized network of 135 Regulatory Discordance Hubs. At these loci, gene expression is dominated by epigenetic state—specifically promoter DNA methylation—rather than underlying copy number alterations. To quantify this phenomenon, we developed the Epigenetic Discordance Score (EDS), which successfully captures a distinct biological state of molecular dosage buffering. In tumors with high epigenetic discordance, the expected proportional relationship between physical DNA dosage and transcript abundance is actively neutralized, demonstrating that epigenetic rewiring can fundamentally reconfigure a tumor's regulatory architecture independently of global genomic instability.

Despite the robust biological signal captured by the EDS, our comprehensive survival analysis reveals that this aggregate state of regulatory rewiring does not translate into a novel axis of clinical lethality. Utilizing a competing risks framework to explicitly isolate disease-specific mortality from non-disease deaths, we found that the EDS does not provide independent prognostic value once established clinical and molecular taxonomies, such as the integrative clusters (INTCLUST), are taken into account. While the aggregate score lacks independent utility, it is notable that several individual regulatory hubs, including *AFF3* and *PGR*, exhibit significant univariate associations with disease-specific death. This suggests that while specific, localized epigenetic rewiring events are linked to tumor progression, the global buffering state itself is not independently lethal.

Ultimately, our findings indicate that the molecular organization of these discordance hubs likely reflects biological states that are already embedded within existing expression-based classifiers. While the epigenetic rewiring of dosage-uncoupled hubs represents a mechanistically significant feature of breast cancer biology that actively buffers physical genomic instability, it does not constitute an independent predictor of patient survival. Future investigations into molecular dosage buffering may benefit from exploring its role in therapeutic resistance or tumor evolution, rather than its direct impact on baseline prognostic stratification.