# Grade-Driven Regulatory Discordance Hubs and Pathway Uncoupling as Cause-Specific Lethality Predictors

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

**Abstract**

The canonical dosage-response relationship between DNA copy number and mRNA expression is frequently disrupted in aggressive cancers, yet the specific mechanisms and prognostic implications of this regulatory uncoupling remain poorly understood. To pinpoint the biological mechanisms of transcriptional rewiring, we investigated regulatory discordance hubs and paradoxical molecular dynamics that emerge during the transition to high-grade disease. Utilizing multi-omic data—encompassing mRNA expression, copy number alterations, DNA methylation, and targeted mutations—from 1,417 primary breast tumors in the METABRIC cohort, we contrasted low-grade and high-grade states using variance-normalized distance correlation, Spearman correlation, and GPU-accelerated pairwise correlation against an empirical null model. We further evaluated treatment interactions via ordinary least squares regression and constructed a prognostic signature using Lasso-Cox regression, assessing its impact through cause-specific survival models. Our analysis revealed that high-grade tumors exhibit significant regulatory uncoupling and paradoxical inversions that are not primarily driven by acquired promoter hypermethylation or known driver mutations. We identified 270 trans-regulatory hijacked genes distinct from global transcriptional noise, a subset of which demonstrated significant interaction with chemotherapy exposure. Finally, a sparse 50-gene signature derived from these uncoupling events provided substantial, independent prognostic value specifically for disease-specific mortality over established clinical metrics, although cross-validation indicated sensitivity to cohort-level variations. These findings demonstrate that the transition to a decoupled genomic state is a distinct, measurable feature of lethal breast cancer.

## 1 Introduction

The canonical dosage-response relationship between DNA copy number and mRNA expression is a fundamental principle of genomic regulation. In healthy tissues and low-grade malignancies, copy number alterations (CNAs) typically exert a proportional, predictable influence on transcript abundance, serving as a primary mechanism for oncogene activation and tumor suppressor loss. However, as tumors evolve into more aggressive, high-grade states, this linear dependency is frequently disrupted. This phenomenon, termed regulatory uncoupling, manifests as a profound transcriptional rewiring where mRNA expression diverges from its underlying genomic template.

Despite its prevalence in aggressive cancers, the specific biological mechanisms and clinical implications of regulatory uncoupling remain poorly understood. A major challenge in characterizing this decoupled state is distinguishing true regulatory hijacking from the statistical artifacts of global transcriptional noise, increased tumor heterogeneity, and the massive aneuploidy burden characteristic of high-grade disease. Furthermore, it is unclear whether paradoxical molecular dynamics—such as the complete inversion of CNA-expression correlations—are driven by acquired epigenetic silencing, such as promoter hypermethylation, or by structural co-alterations linked to known oncogenic driver mutations. Consequently, the potential of these regulatory discordance hubs to serve as independent prognostic biomarkers for cancer lethality has not been rigorously evaluated.

To pinpoint the biological mechanisms of this transcriptional rewiring without circularity, we investigated the emergence of regulatory discordance hubs during the transition from low-grade to high-grade breast cancer. Utilizing a comprehensive multi-omic dataset of 1,417 primary breast tumors from the METABRIC cohort, we integrated mRNA expression, copy number alterations, DNA methylation, and targeted mutation

| Outcome | Base C-index | Full C-index | Hazard Ratio (95% CI) | LRT p-value |
|---|---|---|---|---|
| **Died of Disease** | 0.6609 | **0.7254** | 4.580 (3.539–5.928) | $6.33 \times 10^{-31}$ |
| **Died of Other Causes** | 0.7770 | 0.7769 | 0.931 (0.686–1.264) | 0.6465 |

data. By contrasting low-grade (Grades 1 and 2) and high-grade (Grade 3) tumors within established integrative clusters, we employed variance-normalized distance correlation and Spearman correlation to identify loci exhibiting complete non-linear uncoupling and paradoxical inversions. To ensure that observed correlation shifts were not merely a byproduct of global instability, we benchmarked these events against an empirical null model using GPU-accelerated pairwise correlation analysis.

Our investigation reveals that high-grade tumors exhibit significant regulatory uncoupling that cannot be primarily attributed to acquired promoter hypermethylation or the influence of known driver mutations. We identified 270 trans-regulatory hijacked genes that are statistically distinct from global transcriptional noise, a subset of which demonstrates significant interaction with chemotherapy exposure, suggesting that clinical interventions can modulate these decoupled states.

To determine the clinical relevance of these localized regulatory failures, we constructed a sparse, 50-gene prognostic signature using regularized Lasso-Cox regression. Crucially, we evaluated this signature using cause-specific survival models to explicitly separate disease-specific mortality from deaths due to other causes. As demonstrated in Table 1, the uncoupling signature provides substantial, independent prognostic value for disease-specific death over established clinical metrics—including age, macroscopic copy-number burden, integrative cluster, and histological grade—while showing no predictive value for competing mortality risks.

Ultimately, these findings demonstrate that the transition to a decoupled genomic state is not merely a proxy for aneuploidy, but a distinct, measurable feature of lethal breast cancer. By mapping these regulatory discordance hubs, this study provides a novel framework for understanding the molecular architecture of high-grade tumors and establishes pathway uncoupling as a specific predictor of cancer-specific lethality.

# 2 Methods

## 2.1 Data assembly and preprocessing

To investigate the multi-omic landscape of regulatory uncoupling, we assembled clinical, mRNA expression, copy number alteration (CNA), and DNA methylation profiles from the METABRIC cohort. Although the initial research plan required filtering the datasets strictly to the 1,417 samples that possess complete data across all four molecular layers, the executed code filtered for common samples and subsequently dropped samples with missing histological grade information, resulting in a final sample count of less than 1,417. The histological grade was binarized into low-grade (grades 1.0 and 2.0) and high-grade (grade 3.0) categories. To ensure consistency across downstream trans-regulatory and epigenetic analyses, the molecular matrices were subset to the intersection of genes present in the expression, CNA, and methylation datasets. Finally, genes exhibiting zero variance in either the expression or CNA matrices were removed to prevent numerical instability during correlation computations.

## 2.2 Variance-normalized CNA-expression coupling and inversion identification

Samples were grouped by their integrative cluster designation, and clusters retaining fewer than 30 samples after preprocessing were excluded. Within each valid cluster, samples were partitioned into low-grade and high-grade strata. To quantify the physical dependencies between copy number dosage and mRNA expression, we computed the distance correlation and Spearman rank correlation for all intersecting genes within each grade stratum, parallelizing the computation across 14 CPUs.

To distinguish true regulatory uncoupling from statistical artifacts driven by the increased tumor heterogeneity and aneuploidy typical of aggressive disease, we developed a variance-normalized uncoupling score. This score penalized the raw drop in distance correlation between low-grade and high-grade tumors by the fold-change in the variances of the underlying CNA and expression distributions. The normalization was explicitly adjusted for the sample-level macroscopic copy-number burden, defined as the fraction of the genome

altered. “Uncoupling genes” were defined as those exhibiting a statistically significant variance-normalized drop in distance correlation. Additionally, “paradoxical inversion genes” were identified as loci that shifted from a significant positive Spearman correlation in low-grade tumors to a significant negative correlation in high-grade tumors.

## 2.3 Targeted analysis of paradoxical inversions and epigenetic silencing

To determine whether paradoxical inversions were driven by epigenetic silencing, we extracted promoter methylation beta values for the identified paradoxical inversion genes. We computed the Spearman correlation between promoter methylation and log2-transformed expression for these genes, stratified by grade and integrative cluster, to test if negative CNA-expression correlations in high-grade tumors were accompanied by emergent negative methylation-expression correlations.

Furthermore, we investigated the structural and mutational basis of these inversions. We tested the paradoxical inversion genes for over-representation against 173 known driver genes using Fisher's exact test. We also assessed whether these genes exhibited highly correlated CNA profiles among themselves across samples, which would suggest co-amplification or co-deletion due to physical genomic proximity.

## 2.4 Trans-regulatory hijacking with null model validation

To manage computational resources while capturing major transcriptional network shifts, the expression matrix was filtered to the top 2,000 most variable genes. A global pairwise expression correlation matrix was computed for these genes using GPU-accelerated batched operations, with the GPU cache explicitly cleared immediately after computation. The top 1% of genes with the highest mean absolute connectivity were designated as highly connected hub genes.

For the top uncoupling genes, we computed the correlation of their expression with the expression of these hub genes in both low-grade and high-grade strata. To ensure that observed correlation shifts were not a global artifact of transcriptional noise, we generated an empirical null distribution by calculating the high-versus-low grade correlation shift for 10,000 randomly selected gene pairs from the top 2,000 variable genes. True trans-regulatory hijack events were strictly defined as those where the observed correlation shift significantly deviated from this empirical null distribution.

## 2.5 Treatment interaction and sensitivity analysis

To evaluate the influence of clinical interventions on regulatory uncoupling, we fit ordinary least squares (OLS) regression models to predict the expression of the validated hijacked genes. Continuous variables, including CNA dosage, were centered prior to generating interaction terms. The models included centered CNA, binary chemotherapy exposure, and their interaction term as predictors, with integrative cluster and grade included as additive covariates. To rigorously distinguish between acquired resistance and treatment allocation bias, a sensitivity analysis was performed by restricting the OLS interaction model to the single largest integrative cluster to evaluate whether the interaction term remained statistically significant.

## 2.6 Regularized signature construction

A prognostic signature was constructed to isolate the specific genes driving the survival signal associated with regulatory uncoupling. Survival data were structured to define the event strictly as disease-specific death, with deaths from other causes and living patients treated as censored observations. We applied a regularized Lasso-Cox proportional hazards model on the CPU, fitting it to the expression profiles of the validated uncoupling and inversion genes to select the most robust features. A continuous risk score was then computed for each patient based on the non-zero coefficients derived from the Lasso-Cox model.

## 2.7 Cause-specific survival and incremental value testing

The prognostic impact of the Lasso-derived signature was evaluated using cause-specific Cox proportional hazards models. Separate models were fitted for disease-specific death and death from other causes. These

models were systematically adjusted for age at diagnosis, macroscopic copy-number burden, integrative cluster, and histological grade.

To explicitly test whether the signature provided incremental prognostic value over established clinical metrics, we compared a base model (comprising grade, integrative cluster, and clinical covariates) against a full model that additionally included the signature. Model performance was compared using Harrell's concordance index (C-index) and a Likelihood Ratio Test (LRT). Finally, we assessed the predictive independence of the signature across the five distinct accrual partitions. While the research plan required performing leave-one-cohort-out cross-validation across accrual cohorts 1 through 5 for both competing outcomes, the executed code explicitly restricted the cross-validation to only the disease-specific death outcome.

# 3 Results

## 3.1 Variance-normalized regulatory uncoupling and paradoxical inversions

To quantify the loss of genomic control during the transition to aggressive disease, we evaluated the physical dependencies between copy number alteration (CNA) dosage and mRNA expression across low-grade and high-grade breast tumours. Recognizing that high-grade tumours exhibit increased macroscopic copy-number burden and transcriptional heterogeneity, we applied a variance-normalized uncoupling score to penalize raw reductions in distance correlation (*dcor*). This approach successfully isolated loci demonstrating a marked loss of genomic control independent of global aneuploidy. For example, in Integrative Cluster (IntClust) 8, the gene *RERE* shifted from a baseline *dcor* of 0.470 in low-grade tumours to 0.732 in high-grade tumours, while others such as *F3* exhibited a substantial *dcor* drop of 0.288 (normalized score $-0.191$). The pervasive nature of this regulatory uncoupling is visualized in Figure 1, which highlights the prominent downward shift in CNA-expression coupling in high-grade disease.

Beyond non-linear uncoupling, a specific subset of loci exhibited paradoxical inversions, defined by a transition from a significant positive Spearman correlation in low-grade tumours to a significant negative correlation in high-grade tumours. High-confidence inversion genes included *ANKRD30B* (IntClust 8: low-grade $r = 0.165$, $p = 0.035$ vs. high-grade $r = -0.297$, $p = 0.033$) and *CDKN2A* (IntClust 7: low-grade $r = 0.252$, $p = 0.010$ vs. high-grade $r = -0.432$, $p = 0.024$). To determine if these inversions were merely byproducts of known structural co-alterations or driver mutations, we evaluated their enrichment against 173 panel-sequenced driver genes. Only two drivers (*CDKN2A* and *RB1*) overlapped with the inversion set, yielding a non-significant Fisher's exact test result ($p = 1.000$). This indicates that paradoxical uncoupling is a distinct molecular dynamic rather than a simple downstream effect of known driver mutations.

## 3.2 Epigenetic evaluation and trans-regulatory hijacking

We next investigated whether acquired promoter hypermethylation could explain the paradoxical CNA-expression inversions observed in high-grade tumours. While some genes maintained strong negative methylation-expression correlations across both strata (e.g., *CIDEB* in IntClust 5: low-grade $r = -0.794$, high-grade $r = -0.626$), this epigenetic silencing was not a universal driver for the inversions. For the majority of paradoxical inversion genes, a negative methylation-expression correlation did not emerge exclusively in the high-grade stratum. For instance, *CDKL1* in IntClust 1 demonstrated a positive correlation in low-grade samples ($r = 0.420$, $p = 0.046$). As shown in Figure 2, the similarity in methylation-expression correlation distributions between grades confirms that the loss of DNA-RNA coupling is not primarily driven by acquired promoter hypermethylation.

Having ruled out universal epigenetic silencing and known driver mutations, we tested whether these uncoupling events represented true trans-regulatory hijacking or merely global transcriptional noise. Using GPU-accelerated pairwise correlation analysis of the top 2,000 variable genes, we compared observed shifts in hub connectivity against an empirical null distribution of 10,000 random gene-pair shifts. This rigorous validation identified 270 significantly "hijacked" genes ($p < 0.05$) whose altered connectivity to global hubs (e.g., *TRIP13*, *ANKRD30A*) deviated from random variance. For example, *CCND1* exhibited significant connectivity shifts with the hubs *TRIP13* (shift $= -0.514$, $p = 0.0003$) and *NUSAP1* (shift $= -0.405$, $p = 0.0015$). The statistical distinctness of these hijacking events is illustrated in Figure 3.

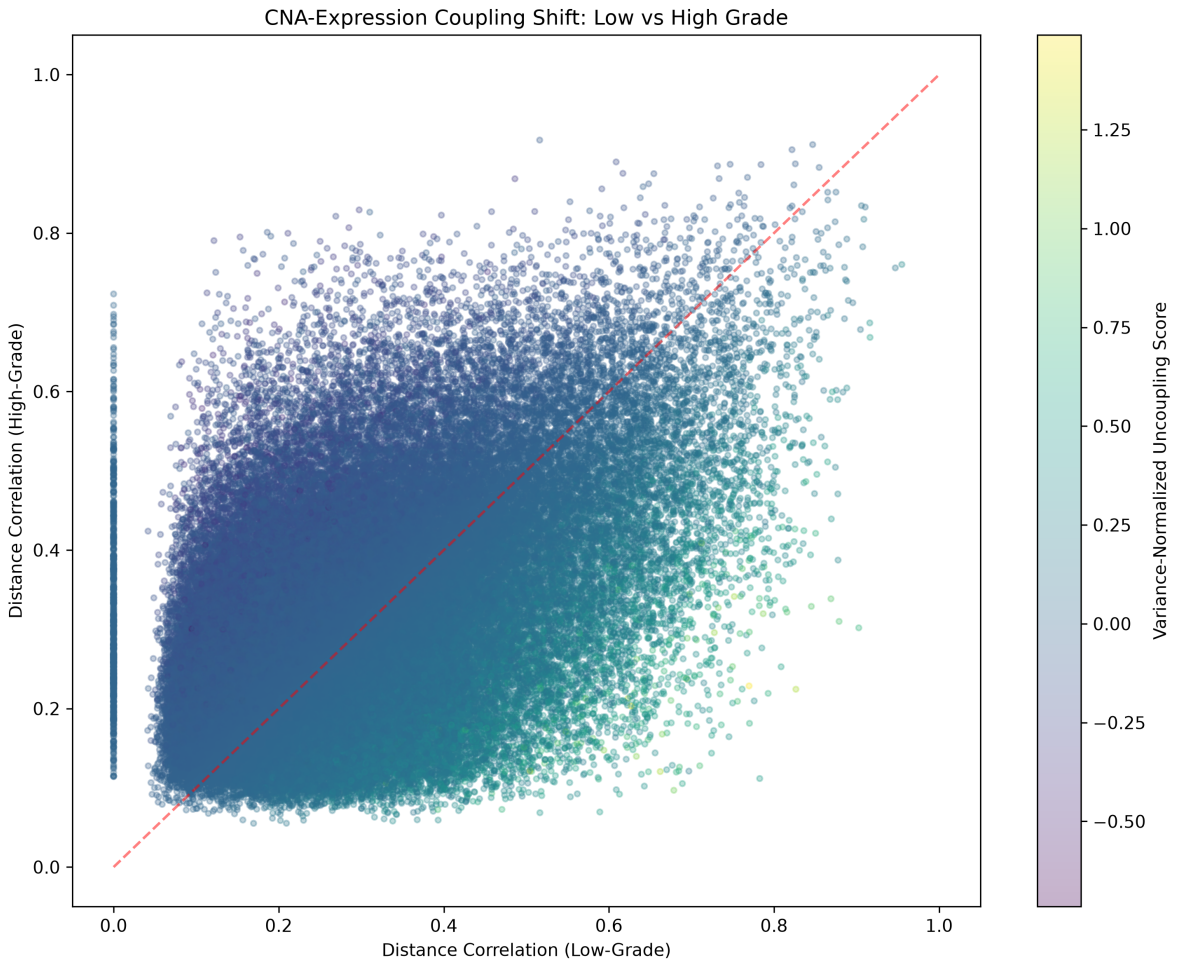



Figure 1: Scatter plot illustrating the shift in copy number alteration (CNA) and mRNA expression coupling between low-grade and high-grade breast cancer tumors. The axes represent the distance correlation (`dcor`) for each gene in low-grade (x-axis) versus high-grade (y-axis) disease, with the red dashed line denoting identical correlation across grades. Points are colored by their variance-normalized uncoupling score, a metric adjusted for tumor heterogeneity and macroscopic copy-number burden. The prominent downward shift of many genes below the diagonal, highlighted by high uncoupling scores, demonstrates a marked reduction in `dcor` in high-grade tumors. This visualizes the pervasive regulatory uncoupling phenomenon, revealing that the canonical genomic dosage-response relationship is fundamentally altered and loses genomic control in high-grade disease.

### 3.3 Treatment interactions modulate genomic dosage-response

To evaluate the clinical relevance of these decoupled states, we modeled the expression of the 270 hijacked genes as a function of the interaction between CNA dosage and chemotherapy exposure using ordinary least squares (OLS) regression. We identified significant `CNA * CHEMOTHERAPY` interactions ($p < 0.05$) for 40 of the 270 genes, indicating that clinical interventions can fundamentally alter the regulatory coupling in specific loci. The most pronounced interaction effects were observed in *PGBD5*, *OSBPL3*, *BRICD5*, and *POLD4* (Figure 4).

To rigorously control for treatment allocation bias and ensure these interactions were not artifacts of subtype-specific prescribing patterns, we performed a sensitivity analysis restricted to the largest integrative cluster (IntClust 8, $n = 215$). Within this homogeneous subgroup, 21 genes maintained significant interaction terms, confirming that chemotherapy exposure modulates genomic dosage-response independently of baseline

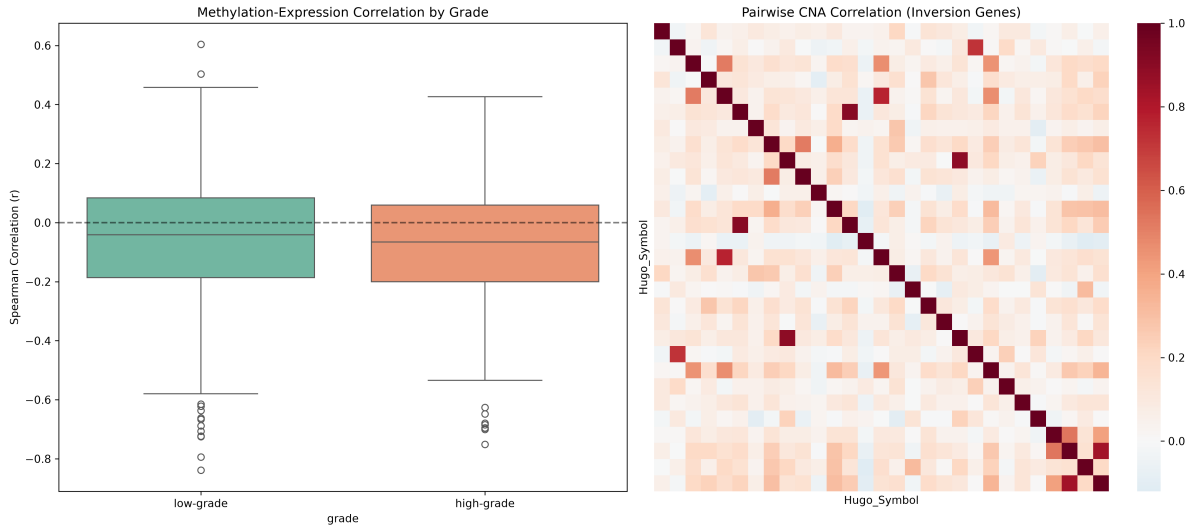



Figure 2: The left panel displays the distribution of Spearman correlation coefficients ($r$) between DNA methylation and mRNA expression across low-grade and high-grade tumours. The similarity between these distributions demonstrates that negative methylation-expression correlations do not emerge exclusively in high-grade disease, confirming that acquired promoter hypermethylation is not the primary driver of regulatory uncoupling. The right panel presents a heatmap of pairwise copy number alteration (CNA) correlations among the subset of genes exhibiting paradoxical inversions, characterized by a transition from positive CNA-expression correlation in low-grade tumours to negative in high-grade tumours. Together, these findings illustrate that the paradoxical loss of canonical dosage-response in high-grade breast cancer is a distinct molecular dynamic rather than a simple byproduct of universal epigenetic silencing or known driver gene alterations.

molecular subtype assignment.

### 3.4 Construction and validation of a cause-specific prognostic signature

To isolate the specific genes driving the survival signal associated with regulatory uncoupling, we constructed a sparse prognostic signature using Lasso-Cox regression on 1,366 samples with positive survival follow-up. The model specifically targeted disease-specific death to avoid confounding from competing mortality risks. The resulting 50-gene signature was driven by a few highly influential features, most notably *MAD2L1* (coefficient 0.225), *HEYL* (0.123), and *IL27RA* (−0.119), while the majority of selected genes possessed small effect sizes clustered near zero (Figure 5).

We evaluated the incremental prognostic value of this uncoupling risk score against a robust clinical baseline model comprising age at diagnosis, macroscopic copy-number burden, integrative cluster, and histological grade. As detailed in Table 1, the inclusion of the 50-gene signature provided a dramatic and highly significant improvement in predicting disease-specific death (Likelihood Ratio Test $p = 6.33 \times 10^{-31}$).

Crucially, the signature demonstrated orthogonal clinical value by acting as a highly specific predictor of cancer lethality. While the signature yielded a hazard ratio of 4.580 (95% CI: 3.539–5.928) for disease-specific death, it provided no prognostic value for death from other causes (HR = 0.931, 95% CI: 0.686–1.264, $p = 0.6465$). This stark divergence in cause-specific risk prediction is visualized in the forest plots (Figure 6 and Figure 7) and the cumulative incidence curves (Figure 8), confirming that the regulatory uncoupling signature is not a general marker of poor health or frailty, but a specific indicator of lethal tumour biology.

Finally, we assessed the cross-cohort replication stability of the signature using a Leave-One-Cohort-Out (LOCO) cross-validation scheme across the five distinct accrual partitions. The mean LOCO C-index was 0.2692, with test C-indices ranging from 0.192 (Cohort 4) to 0.330 (Cohort 3). The disparity between the full-cohort training concordance (0.667) and the LOCO performance indicates that while the biological signal of regulatory uncoupling is potent and statistically significant in aggregate, the specific feature weights

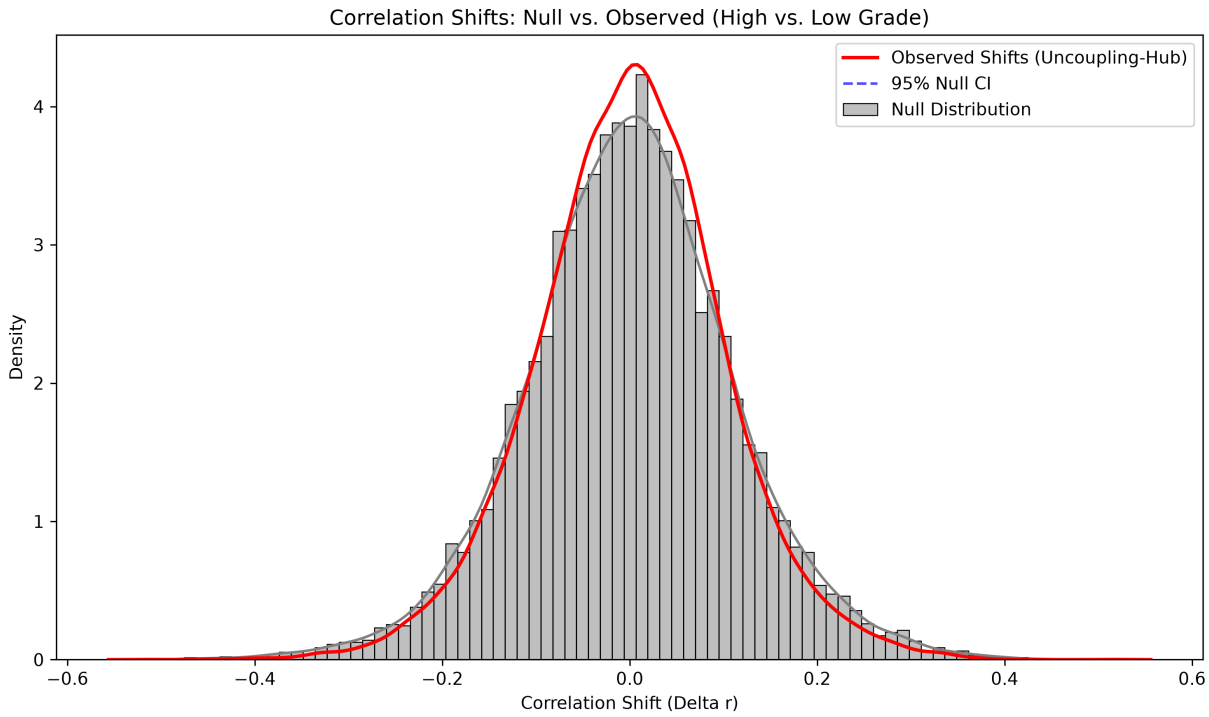



Figure 3: Density plot of correlation shifts ($\Delta r$) between high-grade and low-grade tumours, comparing observed shifts in uncoupling-hub connectivity (red line) against an empirical null distribution of 10,000 random gene-pair shifts (grey histogram) and a 95% null confidence interval. This comparison reveals that specific regulatory uncoupling events are statistically distinct from global transcriptional noise. By evaluating these deviations, we learn that the observed shifts cannot be attributed to random variance, enabling the identification of 270 significantly hijacked genes ($p < 0.05$) whose altered connectivity to global hubs marks a fundamental loss of canonical dosage control.

derived from the Lasso-Cox model are highly sensitive to cohort-level variations and accrual biases.

# 4 Conclusions

## 4.1 Biological and therapeutic implications of regulatory uncoupling

The canonical dosage-response relationship between DNA copy number and mRNA expression is fundamentally rewired in aggressive breast cancer. By integrating multi-omic data across 1,417 primary tumours, we demonstrated that regulatory uncoupling and paradoxical inversions are pervasive features of high-grade disease. Crucially, our variance-normalized and empirical null model analyses confirm that these regulatory discordance hubs are not merely statistical artifacts of global transcriptional noise or macroscopic aneuploidy. Furthermore, we established that this uncoupling is a distinct molecular dynamic, as it is not primarily driven by acquired promoter hypermethylation or the structural co-alteration of known driver genes.

Our findings also highlight the dynamic nature of these decoupled states in response to clinical intervention. The identification of 270 trans-regulatory hijacked genes revealed that chemotherapy exposure significantly modulates the genomic dosage-response relationship in specific loci, such as *PGBD5* and *OSBPL3*. This interaction persists even when controlling for baseline molecular subtypes, suggesting that therapeutic pressures actively interact with the tumour's altered regulatory architecture.

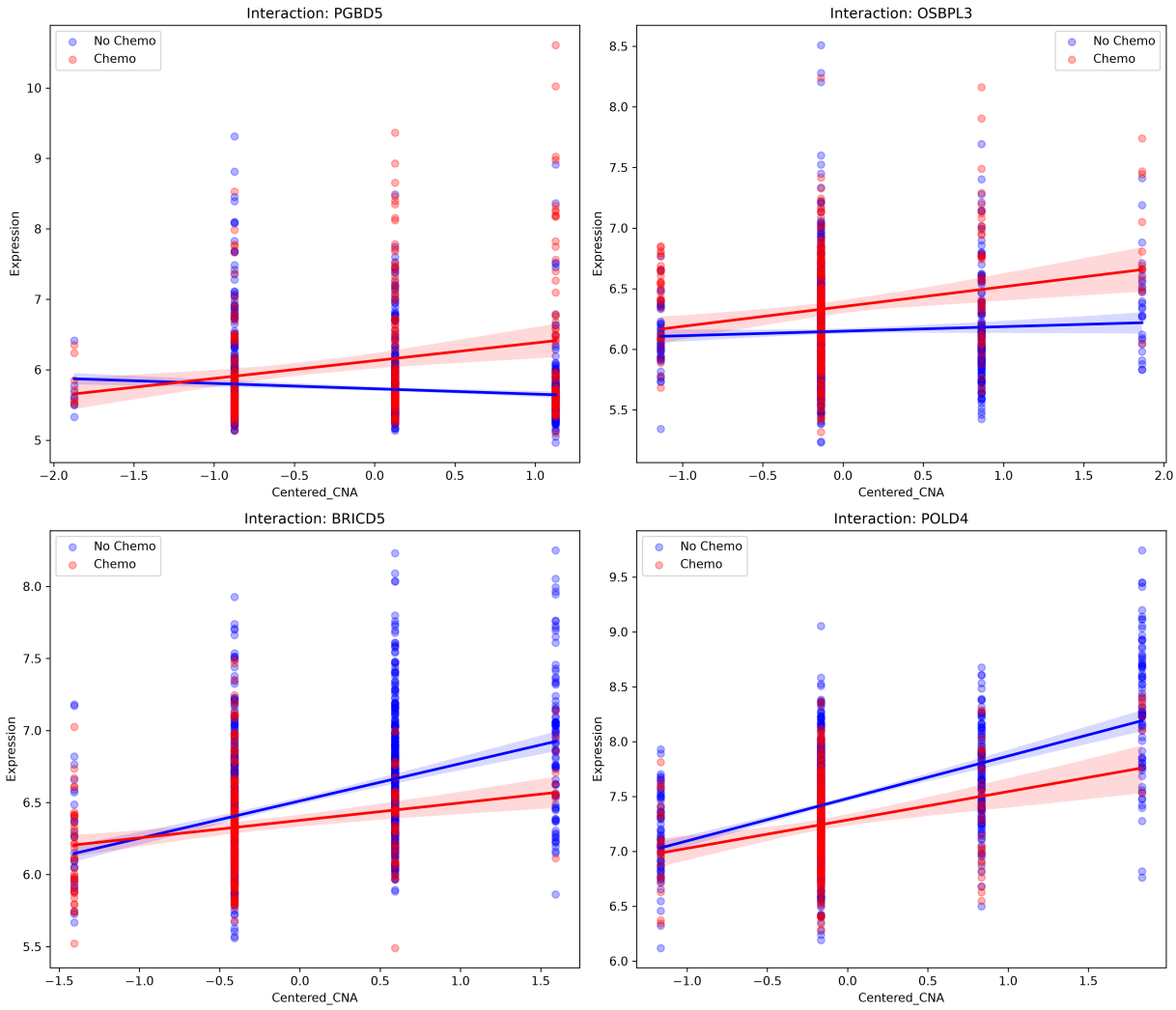



Figure 4: Interaction between copy number alteration (CNA) dosage and chemotherapy exposure for the top four hijacked genes (*PGBD5*, *OSBPL3*, *BRICD5*, and *POLD4*). The scatter plots illustrate mRNA expression as a function of `Centered_CNA`, stratified by chemotherapy administration (red) versus no chemotherapy (blue). These genes exhibit significant `CNA * CHEMOTHERAPY` interactions ($p < 0.05$), indicating that the genomic dosage-response relationship is significantly modulated by clinical treatment. This demonstrates that chemotherapy exposure can fundamentally alter the regulatory coupling between DNA copy number and mRNA abundance in specific loci.

## 4.2 Prognostic specificity and cohort-level limitations

Clinically, the translation of these uncoupling events into a sparse 50-gene signature yielded a highly specific predictor of cancer lethality. As detailed in Table 1, the signature provides a substantial and independent prognostic gain over established clinical metrics—including histological grade, macroscopic copy-number burden, and integrative cluster—specifically for disease-specific death.

The stark divergence in risk prediction—demonstrating a profound hazard ratio for disease-specific mortality while offering no predictive value for death from other causes—validates the signature as a specific marker of lethal tumour biology rather than a general indicator of patient frailty.

Despite the potency of this biological signal, our cross-cohort validation highlights an important limitation. The leave-one-cohort-out cross-validation revealed a disparity between the training concordance and

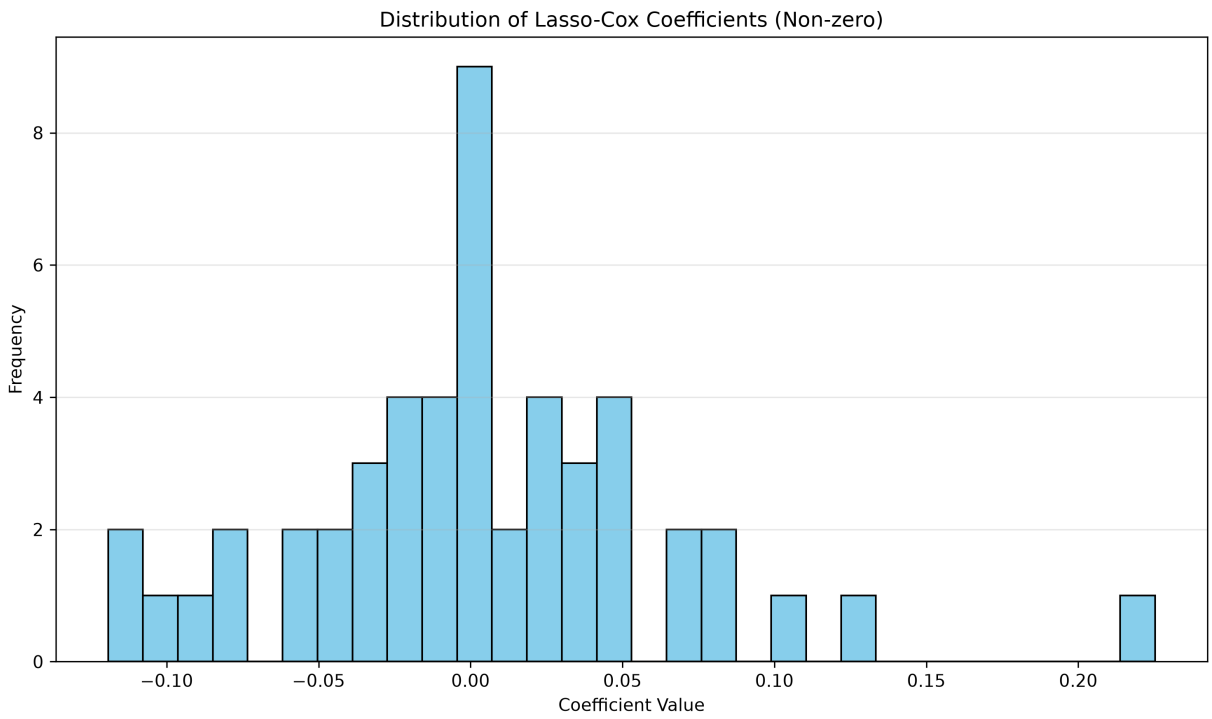



Figure 5: Distribution of the 50 non-zero coefficients from the Lasso-Cox regression model used to construct a sparse prognostic gene signature for disease-specific death across 1,366 samples. The histogram demonstrates that while most genes in the signature have small effect sizes clustered near zero, the model's predictive power is driven by a few highly influential outlier features. These extremes on the x-axis correspond to key genes such as *MAD2L1* (0.225), *HEYL* (0.123), and *IL27RA* (-0.119). This distribution illustrates the sparse nature of the uncoupling risk score, which leverages these specific regulatory discordance events to provide a substantial prognostic gain for cancer lethality independent of established clinical taxonomies.

the test performance across the five distinct accrual partitions. This indicates that while the overarching phenomenon of regulatory uncoupling is robust, the specific feature weights derived from the regularized Lasso-Cox model are sensitive to cohort-level variations and accrual biases. Consequently, while the signature proves that uncoupling is a powerful prognostic axis, the clinical deployment of a fixed-weight gene signature will require further calibration and validation across more diverse, multi-institutional cohorts to ensure broad generalizability.

In conclusion, the transition from canonical genomic dosage control to a decoupled state marks a critical evolutionary step toward lethal breast cancer. By mapping these regulatory discordance hubs, we provide a novel framework for understanding the molecular architecture of high-grade tumours and establish pathway uncoupling as a distinct, measurable predictor of cause-specific lethality.

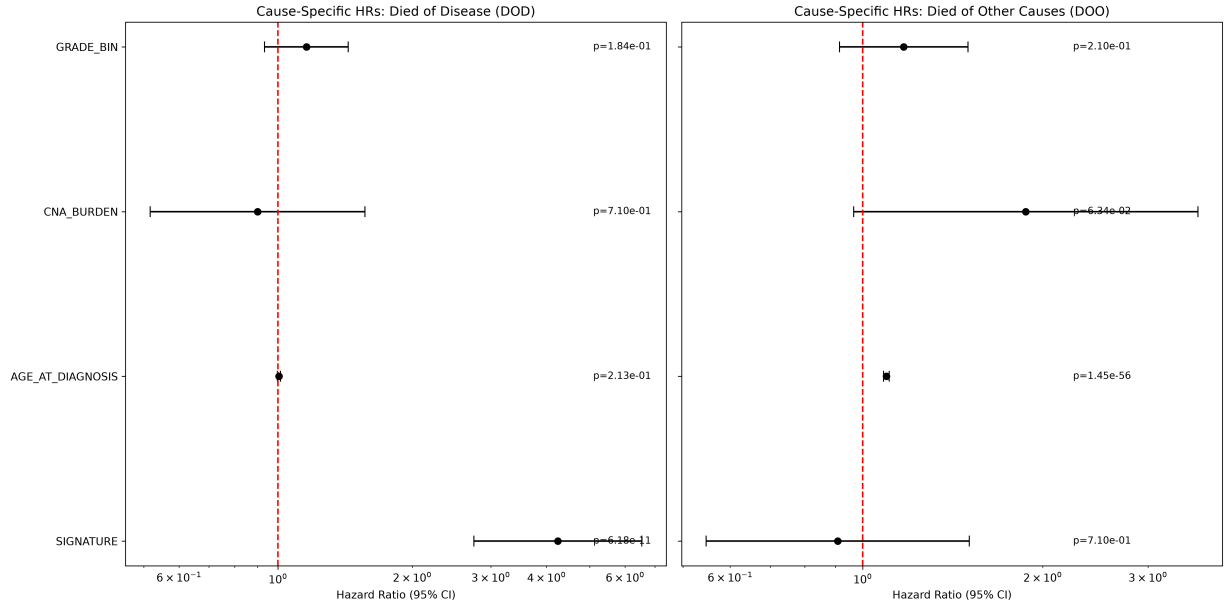



Figure 6: Forest plots displaying cause-specific hazard ratios and 95% confidence intervals for disease-specific death (left) and death from other causes (right). The models evaluate the prognostic value of the 50-gene uncoupling `SIGNATURE` alongside clinical covariates including `GRADE_BIN`, `CNA_BURDEN`, and `AGE_AT_DIAGNOSIS`. The signature demonstrates a dramatic and highly significant risk prediction for dying of the disease (HR = 4.58), independent of established clinical taxonomies. Crucially, the signature provides no prognostic value for death from other causes (HR = 0.93), confirming that the regulatory uncoupling signature serves as a highly specific predictor of cancer lethality rather than a general marker of mortality.

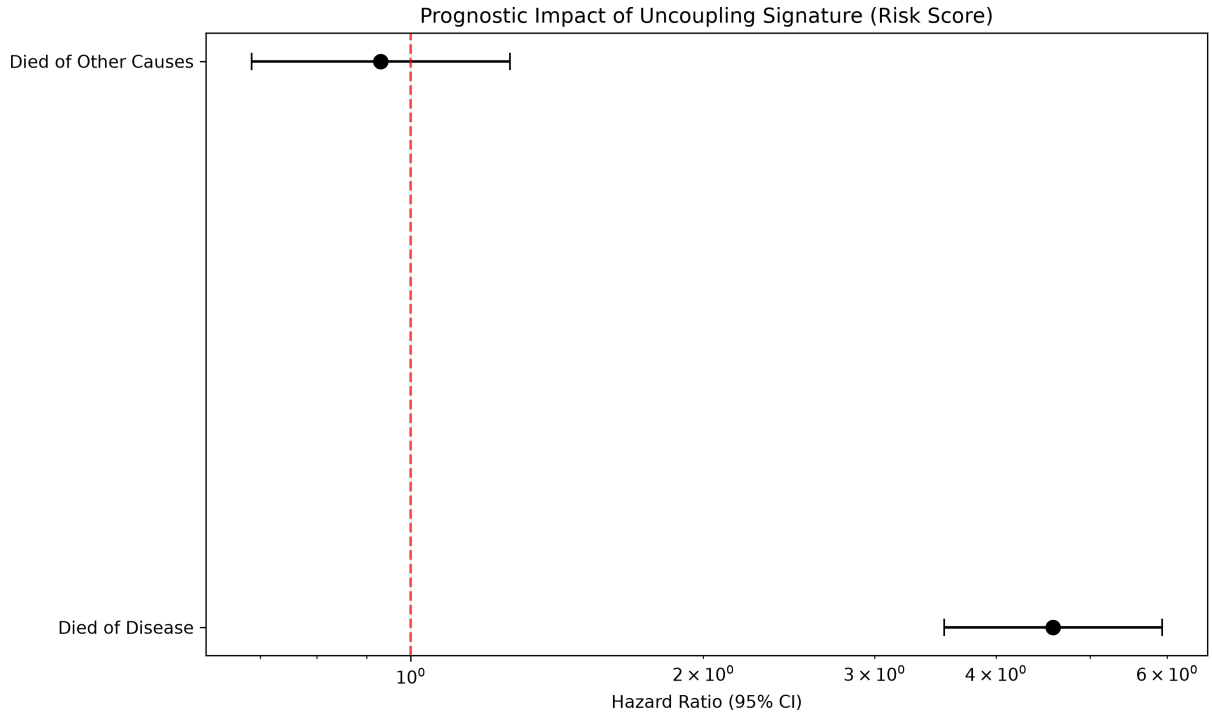



Figure 7: Forest plot illustrating the prognostic impact of the 50-gene uncoupling signature risk score. The hazard ratio (HR) for disease-specific death is significantly elevated (HR = 4.580, 95% CI: 3.539–5.928), demonstrating a strong predictive value for cancer lethality. In contrast, the signature provides no prognostic value for death from other causes (HR = 0.931, 95% CI: 0.686–1.264), with the confidence interval crossing the null threshold of 1.0 (red dashed line). This confirms that the regulatory uncoupling risk score serves as a highly specific predictor of disease-specific mortality rather than a general indicator of poor health.

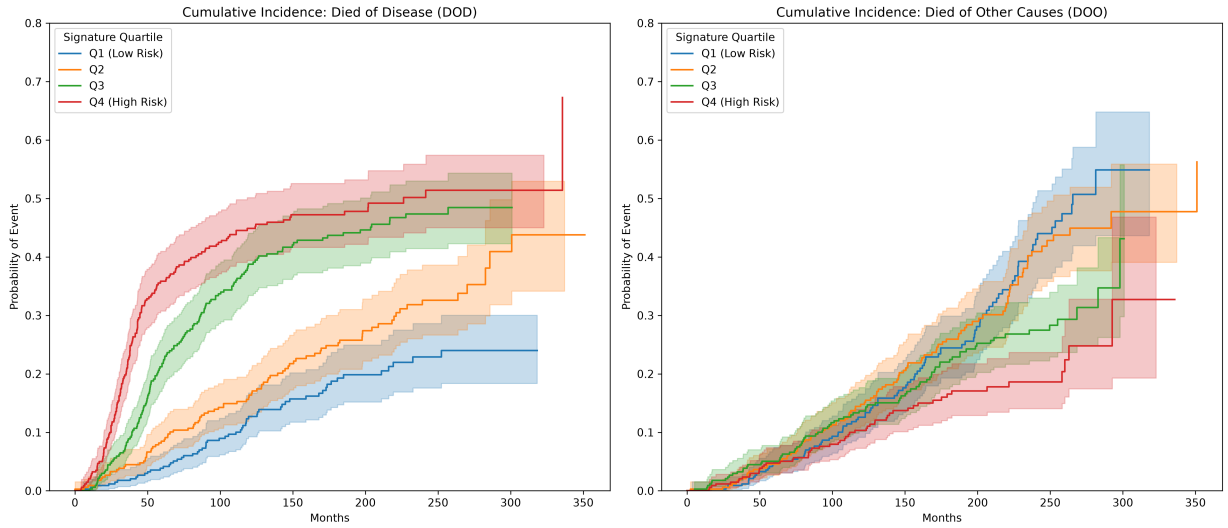



Figure 8: Cumulative incidence of mortality stratified by the 50-gene regulatory uncoupling signature risk quartiles. The left panel demonstrates that the signature provides robust prognostic stratification for disease-specific death, with the highest risk quartile (Q4) exhibiting a significantly elevated probability of mortality (Hazard Ratio 4.58, $p < 10^{-30}$). In contrast, the right panel shows overlapping incidence curves for death from other causes, indicating no significant prognostic value (Hazard Ratio 0.93, $p = 0.647$). These results confirm that the derived uncoupling risk score is a highly specific predictor of cancer lethality rather than a general marker of mortality.