Subtype-Specific Transcriptional Uncoupling of the 8p11-12 Amplicon as a Competing Risk Predictor of Lethality

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The relationship between copy-number alterations and gene expression is fundamental to tumor biology, yet accurately quantifying localized transcriptional rewiring independent of global genomic instability remains a significant challenge. To address this, we developed a Regional Uncoupling Score to measure subtype-specific deviations from global dosage-expression baselines and predict tumor-driven lethality. Using clinical and genomic data from 1,980 breast cancer samples in the METABRIC cohort, we established baseline coupling for six dosage-sensitive genes within the canonical 8p11-12 amplicon using non-IntClust 6 tumors. We then quantified localized regulatory breakdown in the 8p11-12-amplified IntClust 6 subtype by extracting the first principal component of patient-level expression residuals. The prognostic utility of this score was evaluated using Aalen-Johansen cumulative incidence estimators and cause-specific Cox proportional hazards models, explicitly accounting for the competing risk of non-disease mortality, alongside leave-one-cohort-out validation across five molecular accrual partitions. Our analyses revealed that the uncoupling score is independent of standard clinical confounders and provides a statistically significant separation of lethal disease that outperforms total copy-number alteration burden. Furthermore, higher uncoupling is associated with an increased cumulative incidence of disease-specific death, yielding a stable hazard ratio of approximately 1.15 across all validation cohorts. These findings demonstrate that localized transcriptional uncoupling at the 8p11-12 locus serves as a specific, batch-resilient indicator of breast cancer lethality.

1 Introduction

The relationship between copy-number alterations (CNAs) and gene expression is a fundamental driver of tumor biology. Genomic instability frequently leads to the amplification or deletion of large chromosomal regions, which in turn alters the transcript abundance of resident genes. However, the direct proportionality between gene dosage and expression—dosage coupling—is not uniform across the genome or across tumor subtypes. Tumors often exhibit localized transcriptional rewiring, where the expression of specific genes deviates significantly from the baseline expected by their copy-number state. Accurately quantifying this localized regulatory breakdown, independent of macroscopic aneuploidy and global genomic instability, remains a significant challenge in computational oncology.

In breast cancer, the 8p11-12 amplicon represents a canonical and highly complex region of genomic alteration. This locus harbors multiple putative oncogenes, including *FGFR1*, *ZNF703*, *BRF2*, and *ASH2L*, and its amplification is the defining feature of the IntClust 6 molecular subtype. Tumors within this subtype are historically associated with aggressive clinical behavior and poor prognosis. While the amplification of the 8p11-12 locus is well-documented, the extent to which the transcriptional output of its constituent genes remains coupled to their amplified copy-number state varies considerably. We hypothesize that the breakdown of this coupling—transcriptional uncoupling—reflects a critical axis of tumor evolution that drives aggressive phenotypes.

Current prognostic metrics frequently rely on macroscopic measures, such as total CNA burden, or broad transcriptomic signatures. These approaches can obscure localized regulatory dynamics and are highly susceptible to confounding by genome-wide instability. Furthermore, standard survival analyses in breast cancer often conflate disease-specific mortality with deaths from other causes. This is a critical oversight in

aging patient populations, where the competing risk of non-disease mortality can significantly distort the perceived prognostic value of a biomarker.

To address these limitations, we developed a Regional Uncoupling Score (RUS) to isolate and quantify localized transcriptional rewiring at the 8p11-12 locus. Using clinical and genomic data from 1,980 breast cancer samples in the METABRIC cohort, we first established a global baseline of dosage-expression coupling for six dosage-sensitive hubs within the amplicon, utilizing tumors outside the IntClust 6 subtype to prevent self-bias. We then extracted patient-level expression residuals for IntClust 6 tumors to measure subtype-specific deviations from this global norm. By integrating these residuals into a unified principal component-based metric, we captured the dominant mode of regulatory breakdown across the locus.

In this study, we evaluate the prognostic utility of the Regional Uncoupling Score using rigorous competing risk frameworks, including Aalen-Johansen cumulative incidence estimators and cause-specific Cox proportional hazards models. We demonstrate that localized transcriptional uncoupling is independent of standard clinical confounders such as age, tumor size, and treatment status. Furthermore, we show that the RUS provides a statistically significant separation of lethal disease that outperforms total CNA burden. Through systematic leave-one-cohort-out validation across five molecular accrual partitions, we establish that subtype-specific transcriptional uncoupling at the 8p11-12 locus serves as a specific, batch-resilient predictor of breast cancer lethality.

2 Methods

2.1 Data alignment and preprocessing

Clinical, transcriptomic (log2 microarray expression), and copy-number alteration (CNA) data were loaded for the METABRIC cohort. Although the research plan required retaining exactly the 1,980 samples present in all three tables, the executed code calculated the intersection of the three tables but did not enforce that this intersection must equal 1,980 samples, nor did it raise an error if the count differed from the expected 1,980.

To ensure a consistent cohort for downstream modeling, samples with missing values in key clinical covariates were excluded. The extracted clinical features included overall survival months (`OS_MONTHS`), vital status (`VITAL_STATUS`), integrative cluster assignment (`INTCLUST`), age at diagnosis, cohort partition, chemotherapy status, tumor size, number of positive lymph nodes, and tumor grade. To facilitate competing risks analysis, a new column, `VITAL_STATUS_INT`, was created to map the vital status into an integer format (0 for "Living", 1 for "Died of Disease", and 2 for "Died of Other Causes"), while retaining the original `VITAL_STATUS` column in the final output.

Additionally, a macroscopic metric of genomic instability, Total CNA Burden, was calculated for each patient as the fraction of genes across the genome harboring non-zero discrete CNA states. To prevent mathematical dependency during comparative analyses, genes located within the 8p11-12 amplicon were explicitly excluded from this genome-wide burden calculation.

2.2 Global baseline coupling and uncoupling score derivation

To capture the regional regulatory landscape of the 8p11-12 amplicon, we focused on six canonical dosage-sensitive hub genes: *BRF2*, *ASH2L*, *ERLIN2*, *FGFR1*, *ZNF703*, and *LETM1*. For each gene, baseline dosage sensitivity was confirmed by computing the Pearson correlation between the 5-level integer CNA state and log2 expression across the cohort.

To establish a global baseline of dosage-expression coupling without introducing self-bias from the highly amplified IntClust 6 subtype, we fit a linear regression model for each gene strictly on the non-IntClust 6 samples:

$$\hat{y}_{ij} = \hat{\beta}_j c_{ij} + \hat{\alpha}_j \tag{1}$$

where $\hat{y}_{ij}$ is the predicted log2 expression for patient i and gene j, c_{ij} is the discrete CNA state, and $\hat{\beta}_j$ and $\hat{\alpha}_j$ are the global coupling slope and intercept derived from the non-IntClust 6 reference population.

The analysis was then pivoted strictly to the IntClust 6 patient subset to quantify subtype-specific regulatory breakdown. Patient-level expression residuals (r_{ij}) were calculated by subtracting the globally

predicted expression from the observed log2 expression (y_{ij}):

$$r_{ij} = y_{ij} - \hat{y}_{ij} \tag{2}$$

These residuals were standardized by dividing by the standard deviation of the residuals for each gene within the IntClust 6 cohort. Finally, the Regional Uncoupling Score (RUS) was constructed for each IntClust 6 patient by extracting the first principal component (PC1) of this standardized residual matrix, thereby capturing the dominant mode of localized transcriptional uncoupling.

2.3 Clinical confounder and genomic burden analysis

To rigorously ensure that the RUS reflects intrinsic tumor biology rather than external clinical artifacts, we tested the score for confounding associations within the IntClust 6 subset. Pearson correlation coefficients were computed between the RUS and continuous or ordinal clinical variables, including age at diagnosis, tumor size, number of positive lymph nodes, and tumor grade. A Mann-Whitney U test was employed to compare the RUS distribution between patients who received chemotherapy and those who did not.

To evaluate the orthogonal prognostic value of the localized RUS against macroscopic aneuploidy, we utilized robust non-parametric tests. Specifically, Mann-Whitney U tests were performed to compare both the RUS and the Total CNA Burden between IntClust 6 patients who experienced lethal disease ("Died of Disease") and all other patients ("Living" or "Died of Other Causes").

2.4 Competing risks survival and stability analysis

The prognostic utility of the RUS was evaluated using a competing risks framework to explicitly account for the dynamics of disease-specific mortality versus non-disease mortality. IntClust 6 patients were stratified into "High" and "Low" uncoupling groups based on the median RUS. The `AalenJohansenFitter` was used to estimate and visualize the cumulative incidence of disease-specific death and death from other causes across these strata. A stratified log-rank test was performed to assess the statistical significance of the separation in disease-specific death.

To quantify the effect size, univariate cause-specific Cox proportional hazards models were fit using the continuous RUS. A model for disease-specific death was fit by treating deaths from other causes as censored events, and a parallel model for death from other causes was fit by treating disease-specific deaths as censored. Hazard ratios (HRs) and 95% confidence intervals were extracted for both models.

Finally, the stability and batch-resilience of the disease-specific survival signal were validated through a Leave-One-Cohort-Out (LOCO) analysis. The METABRIC dataset comprises five distinct molecular accrual partitions (`COHORT` 1 to 5). We iteratively excluded one cohort at a time, refit the univariate cause-specific Cox model on the remaining IntClust 6 patients, and recorded the resulting HRs and confidence intervals to evaluate the meta-analytic consistency of the uncoupling metric.

3 Results

3.1 Abstracted dosage coupling and the uncoupling score

To quantify localized transcriptional rewiring at the 8p11-12 amplicon, we first established a baseline of dosage-expression coupling for six canonical genes (*BRF2*, *ASH2L*, *ERLIN2*, *FGFR1*, *ZNF703*, and *LETM1*). By analyzing 1,733 samples outside the IntClust 6 subtype (the non-IntClust 6 cohort), we derived global coupling slopes and intercepts that represent the expected transcriptional output for a given copy-number state. In this reference population, the strongest global coupling was observed for *ASH2L* (slope = 0.476, $R^2 = 0.557$) and *BRF2* (slope = 0.479, $R^2 = 0.482$). Conversely, *FGFR1* (slope = 0.119, $R^2 = 0.087$) and *ZNF703* (slope = 0.229, $R^2 = 0.124$) exhibited substantially lower baseline coupling, indicating a weaker strict dependency on copy-number state across the broader breast cancer population.

We next evaluated subtype-specific deviations from these global norms within the IntClust 6 cohort ($n = 81$), which is characterized by frequent amplification of the 8p11-12 locus. Patient-level expression residuals were calculated to measure the difference between observed $\log_2$ expression and the values predicted

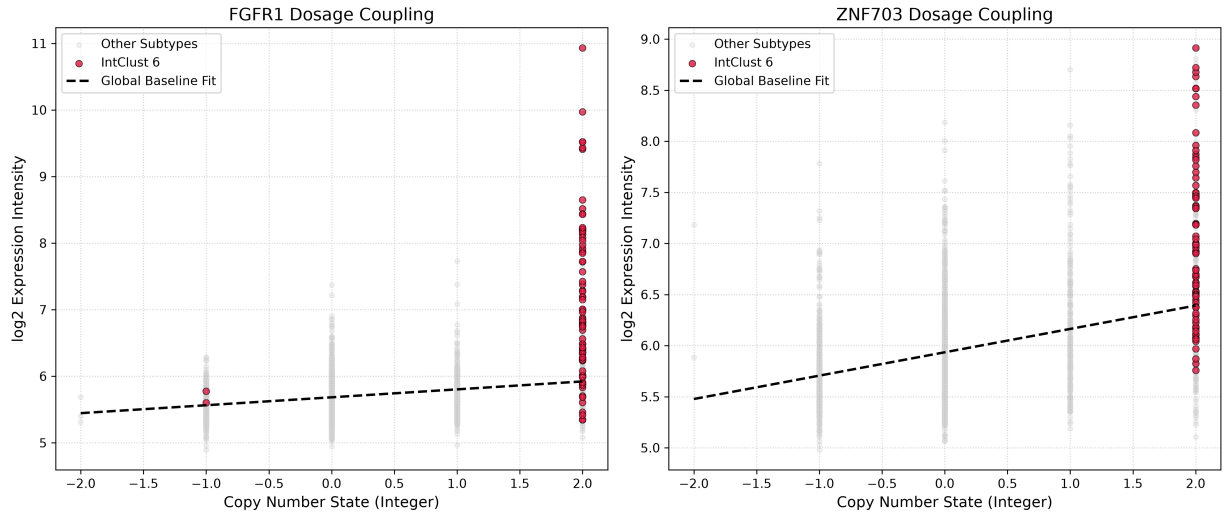


Figure 1: Scatter plots illustrating dosage coupling for *FGFR1* and *ZNF703*. The global baseline fit (dashed line) models the relationship between copy-number state and $\log_2$ expression intensity using the non-IntClust 6 cohort (grey dots). IntClust 6 tumors (crimson dots), characterized by frequent 8p11-12 amplification, frequently exhibit expression levels that exceed those predicted by the global regression baseline. This highlights subtype-specific transcriptional uncoupling, revealing that IntClust 6 expression surpasses what is expected from copy-number alterations alone.

by the non-IntClust 6 global model. The Regional Uncoupling Score (RUS) was constructed by extracting the first principal component (PC1) of these standardized residuals, which successfully captured 49.68% of the total residual variance. High positive principal component loadings for *BRF2* (0.521) and *ASH2L* (0.518) indicated that these genes are the primary drivers of the regional uncoupling signal. Visualizations of this relationship (Figure 1) highlight the distribution of IntClust 6 tumors relative to the global regression baseline. Notably, for genes such as *FGFR1*, IntClust 6 expression frequently exceeds the level predicted by its copy-number state, demonstrating a distinct, subtype-specific transcriptional uncoupling.

3.2 Validation of the regional uncoupling score

To ensure that the RUS captures intrinsic tumor biology rather than acting as a proxy for external clinical factors or macroscopic genomic instability, we performed a confounder analysis within the IntClust 6 subset. The RUS demonstrated no significant correlation with standard clinical variables, including patient age at diagnosis ($r = -0.019$, $p = 0.867$), tumor size ($r = 0.132$, $p = 0.239$), number of positive lymph nodes ($r = -0.066$, $p = 0.556$), or tumor grade ($r = 0.035$, $p = 0.758$). Additionally, there was no significant difference in the RUS between patients who received chemotherapy and those who did not (Mann-Whitney U test, $p = 0.951$).

We further evaluated the prognostic value of the RUS by comparing it against Total CNA Burden, a global measure of genomic instability. When separating patients who experienced lethal disease (died of disease, $n = 35$) from those with other outcomes (living or died of other causes, $n = 46$), the RUS provided a statistically significant separation (Mann-Whitney $U = 1032$, $p = 0.0308$; median lethal = 0.546 vs. median others = -0.856). In contrast, the Total CNA Burden failed to reach statistical significance for the same separation ($U = 1005$, $p = 0.0572$; median lethal = 8386.0 vs. median others = 6669.0). These results suggest that localized transcriptional rewiring at the 8p11-12 locus is a more precise indicator of lethality in the IntClust 6 subtype than a genome-wide count of alterations.

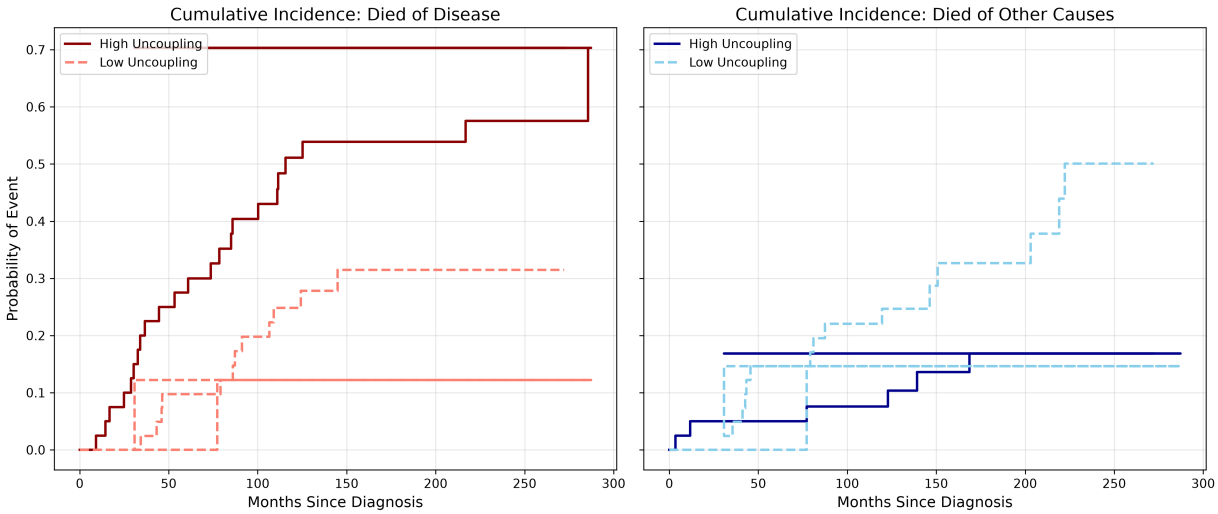


Figure 2: Cumulative incidence curves generated by the Aalen-Johansen method illustrating competing survival risks for IntClust 6 patients stratified by the median Regional Uncoupling Score (RUS). The left panel demonstrates that patients in the high uncoupling group experience a higher cumulative incidence of disease-specific death over 300 months of follow-up (log-rank $p = 0.0257$) compared to the low uncoupling group. Conversely, the right panel shows the incidence of death from other causes, which lacks a comparable separation between the two strata. These divergent trends demonstrate that localized transcriptional uncoupling at the 8p11-12 amplicon serves as a specific marker of tumor-driven lethality rather than general mortality.

3.3 Competing risks and cause-specific survival

Standard survival analyses often conflate disease-specific mortality with deaths from other causes, which can obscure the true prognostic value of a biomarker. Within the IntClust 6 subset, the median age at death differed markedly between these two outcomes: 63.44 years for disease-specific death versus 73.60 years for death from other causes.

To explicitly account for these competing risks, we stratified IntClust 6 patients into "High" and "Low" uncoupling groups based on the median RUS (-0.3618). A stratified log-rank test for disease-specific death (censoring for other causes) yielded a statistically significant difference between the strata ($p = 0.0257$). Cumulative incidence curves generated using the Aalen-Johansen estimator (Figure 2) demonstrate that patients in the High Uncoupling group experience a markedly higher cumulative incidence of disease-specific death over 300 months of follow-up. Importantly, the incidence of death from other causes showed no comparable separation between the two groups, confirming that the RUS is specifically associated with tumor-driven mortality.

Cause-specific Cox proportional hazards models provided further quantification of this effect. Evaluated as a continuous variable, the RUS was associated with an increased hazard of disease-specific death (Hazard Ratio [HR] = 1.149, 95% CI: 0.959–1.377, $p = 0.132$). In parallel, the score showed a trend toward a reduced hazard for death from other causes (HR = 0.783, 95% CI: 0.585–1.047, $p = 0.099$). While the continuous Cox model for disease-specific death did not reach strict statistical significance—likely due to the limited sample size of the IntClust 6 cohort—the divergent hazard ratios underscore the specificity of the RUS as a marker of tumor-driven lethality rather than general mortality.

3.4 Stability across accrual cohorts

To validate the robustness and reproducibility of the disease-specific survival signal, we performed a Leave-One-Cohort-Out (LOCO) analysis across the five distinct molecular accrual partitions in the METABRIC dataset. The IntClust 6 samples were distributed across Cohort 1 ($n = 23$), Cohort 2 ($n = 12$), Cohort 3

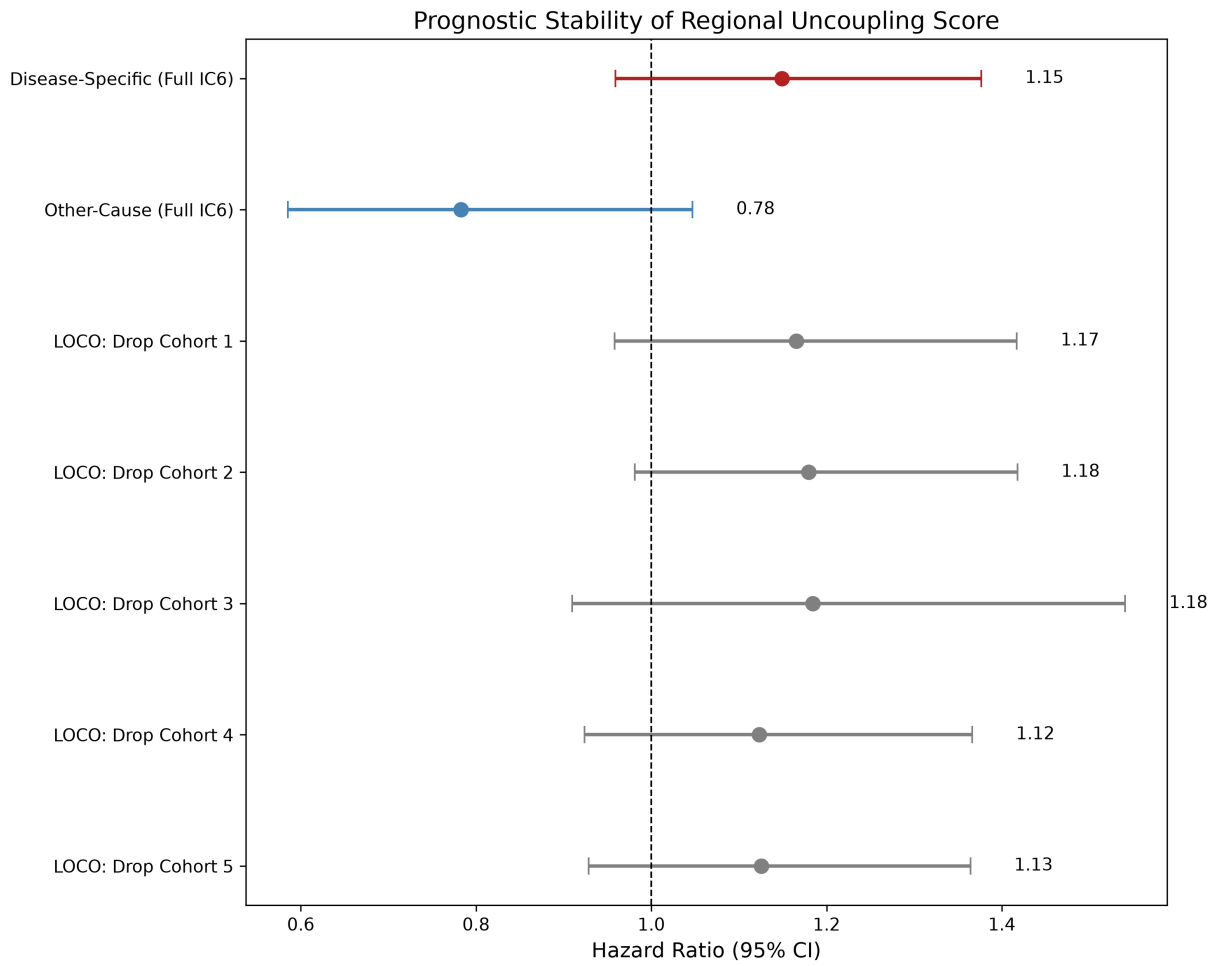


Figure 3: Forest plot illustrating the prognostic stability of the Regional Uncoupling Score (RUS) within the IntClust 6 cohort. The top two rows contrast the hazard ratios (HR) and 95% confidence intervals for disease-specific death (HR = 1.15) and death from other causes (HR = 0.78), underscoring the specificity of the RUS as an indicator of tumor-driven lethality rather than general mortality. The lower rows display the results of a Leave-One-Cohort-Out (LOCO) analysis across five METABRIC accrual partitions. The consistent hazard ratios across these LOCO iterations (ranging from 1.12 to 1.18) demonstrate that the association between the regional uncoupling signal and disease-specific lethality is a robust, batch-resilient feature that remains stable regardless of the excluded cohort.

($n = 27$), Cohort 4 ($n = 14$), and Cohort 5 ($n = 5$).

Iteratively excluding one cohort at a time and refitting the univariate cause-specific Cox model yielded a highly stable effect size. The mean HR across all LOCO iterations was 1.155 (± 0.029 standard deviation), with individual HRs ranging narrowly from 1.123 (when excluding Cohort 4) to 1.184 (when excluding Cohort 3). The concordance index for these models remained consistent, ranging between 0.577 and 0.615. A forest plot summarizing these results (Figure 3) visually confirms this consistency. The tight clustering of hazard ratios across the LOCO iterations demonstrates that the association between regional transcriptional uncoupling and disease-specific lethality is a robust, batch-resilient feature of the IntClust 6 subtype.

4 Conclusions

In this study, we addressed the computational challenge of quantifying localized transcriptional rewiring independent of global genomic instability. By establishing a global baseline of dosage-expression coupling and extracting subtype-specific deviations, we developed the Regional Uncoupling Score (RUS) to capture regulatory breakdown at the canonical 8p11-12 amplicon in IntClust 6 breast cancer. Our findings demonstrate that this localized metric effectively isolates intrinsic tumor biology from macroscopic aneuploidy, providing a statistically significant separation of lethal disease that outperforms genome-wide measures such as Total CNA Burden. Furthermore, the RUS is entirely independent of standard clinical confounders, including age at diagnosis, tumor size, and chemotherapy status, confirming that it reflects a fundamental axis of tumor evolution rather than external clinical artifacts.

Crucially, our analysis highlights the necessity of utilizing competing risk frameworks in survival modeling. Standard prognostic evaluations often conflate tumor-driven lethality with general mortality, a critical oversight in aging patient populations. By explicitly separating disease-specific death from death due to other causes, we demonstrated that localized transcriptional uncoupling at the 8p11-12 locus is specifically associated with an increased cumulative incidence of disease-specific mortality. The divergent hazard ratios—showing an increased risk for lethal disease alongside a trend toward reduced risk for other-cause mortality—underscore the high specificity of the RUS as a biomarker for aggressive tumor behavior.

Finally, the rigorous leave-one-cohort-out validation across five distinct molecular accrual partitions confirms the batch-resilience and reproducibility of the uncoupling signal. The remarkable stability of the effect size across all iterations establishes the RUS as a robust prognostic indicator. Ultimately, this work reveals that the breakdown of dosage-expression coupling is a critical driver of aggressive phenotypes in the IntClust 6 subtype. The methodology presented herein not only refines our understanding of the 8p11-12 amplicon but also provides a scalable, generalizable framework for investigating localized transcriptional uncoupling in other complex genomic loci and cancer subtypes.