Multi-Omic Discordant Hubs as Cause-Specific Predictors of Breast Cancer Lethality

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Breast cancer genomics is often dominated by macroscopic alterations, which can obscure localized multi-omic regulatory failures that drive tumor lethality independently of general patient frailty. To address this, we sought to isolate genuine multi-omic uncoupling by identifying discordant hubs—genes that lack the typical correspondence between DNA copy number, promoter methylation, and mRNA expression. Utilizing a strictly complete four-layer multi-omic dataset of 1,417 breast cancer samples, we applied non-linear modeling to extract a patient-specific regulatory discordance burden. Principal component analysis on subtype-rewired hubs yielded orthogonal discordance factors, which we evaluated using cause-specific Cox proportional hazards models and Aalen-Johansen cumulative incidence functions to account for competing mortality risks. We identified 500 discordant hubs, prominently featuring cell-cycle regulators, that exhibit subtype-specific rewiring. The derived discordance factors were polarized by key driver mutations, such as TP53 and GATA3, but remained independent of global mutational burden. Crucially, incorporating these factors significantly improved the prediction of disease-specific mortality while showing no association with other-cause mortality, establishing the primary discordance factor as a robust, independent predictor of cancer lethality. Finally, the principal axes of discordance demonstrated high structural reproducibility across independent patient cohorts, confirming that localized breakdowns in transcriptional control serve as specific, reproducible markers of lethal tumor biology rather than general genomic instability.

1 Introduction

Breast cancer is a highly heterogeneous disease driven by a complex interplay of genomic, epigenomic, and transcriptomic alterations. Historically, molecular profiling and prognostic modeling have been heavily influenced by macroscopic genomic features, such as global copy-number burden and total mutational burden, alongside established transcriptomic taxonomies. While these large-scale alterations provide valuable prognostic information, they can obscure finer, localized regulatory failures. In a healthy cellular state, mRNA expression is tightly regulated by its genomic and epigenetic scaffolds, with DNA copy number and promoter methylation acting as primary determinants of transcript abundance. However, in the context of malignancy, this regulatory coupling can break down. We hypothesize that localized multi-omic uncoupling—where specific genes lose their typical correspondence with DNA copy number and methylation—represents a distinct biological mechanism of tumor aggressiveness that operates independently of macroscopic genomic instability.

To capture this phenomenon, we introduce the concept of "discordant hubs": genes that exhibit a profound lack of baseline regulatory coupling across multiple molecular layers. When the transcriptome uncouples from its regulatory anchors at these specific nodes, it may enable the tumor to bypass normal cellular checkpoints, particularly those governing the cell cycle. Quantifying this uncoupling provides a patient-specific Regulatory Discordance Burden (RDB). However, a critical challenge in evaluating the clinical utility of any new molecular signature is distinguishing genuine tumor lethality from general patient frailty. Many prognostic markers are evaluated against overall survival, which conflates disease-specific mortality with deaths from other causes. To rigorously establish that multi-omic discordance drives lethal tumor biology rather than reflecting general aging or comorbidities, its prognostic value must be evaluated using a competing risks framework.

In this study, we sought to isolate genuine multi-omic regulatory failure and evaluate its specific impact on breast cancer lethality. Utilizing a strictly complete, four-layer multi-omic dataset of 1,417 breast cancer samples, we systematically identified discordant hubs by measuring the baseline coupling of copy number and methylation to gene expression. By applying non-linear modeling to these hubs, we extracted orthogonal discordance factors that capture subtype-specific regulatory rewiring.

We demonstrate that these discordance factors are polarized by key driver mutations, such as those in *TP53* and *GATA3*, yet remain independent of global mutational burden. Crucially, through the use of cause-specific Cox proportional hazards models and Aalen-Johansen cumulative incidence functions, we show that the primary axes of multi-omic discordance significantly improve the prediction of disease-specific mortality while showing no association with other-cause mortality. Finally, we validate the structural reproducibility of these principal discordance axes across independent patient cohorts. Together, these findings establish localized breakdowns in transcriptional control as specific, reproducible, and orthogonal predictors of breast cancer lethality.

2 Methods

2.1 Data ingestion and cohort alignment

Clinical, transcriptomic (expression), copy number alteration (CNA), DNA methylation, and somatic mutation data were loaded and aligned by patient sample identifiers. While the research plan required intersecting the samples across expression, CNA, methylation, and clinical data to exactly 1,417 samples, the executed code performed an intersection based on the available data in the provided files without enforcing this specific count as a requirement. To adjust for macroscopic genomic instability in downstream analyses, a global copy-number burden was computed for each patient as the fraction of genes in the CNA matrix with a non-zero value. Total mutational burden was calculated as the total number of mutations present within the targeted 173-gene panel. For survival analyses, the vital status was parsed into two mutually exclusive binary event indicators to accommodate competing risks: disease-specific death and death from other causes, utilizing overall survival months as the time-to-event variable. Missing values in the three-gene classifier annotations were explicitly encoded as an "Unknown" category to retain these samples informatively during one-hot encoding.

2.2 Baseline coupling and discordant hub identification

To ensure that the analysis captured biologically active genes, the expression data was first filtered to retain only the top 50% most variable genes across the cohort. For each gene present in the intersection of the expression, CNA, and methylation matrices, baseline regulatory coupling was quantified using the Spearman rank correlation (ρ) between CNA and expression, as well as between methylation and expression.

The cohort median of the absolute correlation values was calculated independently for the CNA-expression and methylation-expression relationships. A candidate pool of weakly coupled genes was defined by isolating genes where both absolute correlations fell below their respective cohort medians. From this pool, the 500 genes with the lowest combined absolute coupling—defined as the minimum sum of squared Spearman correlations—were selected. These genes were designated as empirical "discordant hubs," representing transcriptomic nodes that operate independently of their primary genomic and epigenetic scaffolds.

2.3 Non-linear discordance modeling and factor extraction

To quantify the patient-specific Regulatory Discordance Burden (RDB), the relationship between gene expression and its regulatory anchors was modeled for each of the 500 discordant hubs using an Ordinary Least Squares (OLS) framework. To account for non-linear epigenetic effects and discrete genomic states, the model was specified as:

$$\text{Expression} \sim C(\text{CNA}) + \text{cr}(\text{Methylation}, \text{df} = 3) + C(\text{CNA}) : \text{Methylation}$$

where $C(\text{CNA})$ denotes the categorical treatment of copy number states, and $\text{cr}(\text{Methylation}, \text{df} = 3)$ represents a natural cubic spline with three degrees of freedom applied to the methylation beta-values. Signed

residuals from these models were extracted and standardized (Z-scored) per gene, yielding a matrix of continuous discordance scores for each patient.

To identify hubs exhibiting subtype-specific regulatory rewiring, patients were grouped by their Integrative Cluster (*INTCLUST*) assignments, excluding any clusters with fewer than 50 samples to maintain statistical stability. The median standardized residual was computed for each hub within each valid cluster. The Interquartile Range (IQR) of these median values across clusters was then calculated as a robust measure of cross-subtype dispersion. The top 50 hubs with the highest IQR were designated as "subtype-rewired hubs." Principal Component Analysis (PCA) was performed on this 50-hub matrix, and the first three principal components were extracted as orthogonal Discordance Factors (DF1, DF2, and DF3).

2.4 Mutational landscape analysis

To determine whether multi-omic discordance is driven by specific driver mutations or merely proxies general genomic instability, the top 10 most frequently mutated genes in the cohort (including key regulators such as *TP53* and *GATA3*) were extracted. Binary mutation indicators were generated for each gene. The distributions of the Discordance Factors were compared between mutated and wild-type groups using Mann-Whitney U tests. Additionally, the Spearman correlation between the Discordance Factors and the total mutational burden was computed to assess independence from global mutational rates.

2.5 Cause-specific survival and competing risks modeling

The prognostic utility of the Discordance Factors was evaluated using cause-specific Cox proportional hazards models (with a baseline penalizer of 0.1) to explicitly separate disease-specific mortality from other-cause mortality. For each cause of death, a base clinical model was fit using standardized age at diagnosis, global copy-number burden, total mutational burden, tumor size, number of positive lymph nodes, and one-hot encoded *INTCLUST* and three-gene classifier statuses. A full model was subsequently fit by adding DF1, DF2, and DF3 as continuous main effects. The orthogonal prognostic value of the Discordance Factors was quantified by reporting the change in the Concordance index (C-index) and the Likelihood Ratio Test (LRT) statistic between the base and full models.

To rigorously validate the specificity of the discordance signature for tumor lethality under competing risks, the Aalen-Johansen estimator was utilized to calculate the Cumulative Incidence Function (CIF). The cohort was stratified into tertiles (Low, Mid, High) based on the DF2 score. The CIF and corresponding 95% confidence intervals were estimated and visualized for both disease-specific death and death from other causes across these strata. While cause-specific Cox models were used for hazard ratio and effect size estimation, the Aalen-Johansen estimator was employed to provide the correct non-parametric visualization of cumulative incidence in the presence of competing mortality risks.

2.6 Cohort reproducibility and structural stability

A sensitivity analysis was conducted to verify that the principal axes of discordance represent stable biological structures rather than cohort-specific artifacts. The patient cohort was subset into two independent molecular accrual partitions (Cohort 1 and Cohort 3). PCA was re-run independently on the 50 subtype-rewired hubs within each subset to extract cohort-specific gene loadings. The structural reproducibility of the Discordance Factors was evaluated by calculating the Pearson correlation between the gene weights of Cohort 1 and Cohort 3, matching the principal components by their maximum absolute correlation.

3 Results

3.1 Multi-omic regulatory coupling and identification of discordant hubs

To isolate localized multi-omic uncoupling from macroscopic genomic alterations, we first evaluated the baseline regulatory coupling across the strictly complete four-layer intersection of 1,417 breast cancer samples. Cohort-wide baseline coupling was found to be modest, reflecting a global state where transcript levels are frequently buffered from genomic and epigenetic shifts. The median absolute Spearman correlation

was 0.174 for copy number-expression (CNA-Exp) and 0.046 for methylation-expression (Meth-Exp). By filtering for genes with the lowest combined absolute coupling, we established a set of 500 "discordant hubs." Representative hubs included *PMEPA1* (CNA-Exp $\rho = -0.0004$, Meth-Exp $\rho = 0.0003$), *C1R*, and *ACTB*. These genes represent transcriptomic nodes that operate nearly independently of their primary DNA and epigenetic scaffolds in the cohort.

3.2 Subtype-specific rewiring and extraction of discordance factors

We next applied non-linear modeling to quantify sample-specific deviations from expected regulatory patterns, yielding a continuous Regulatory Discordance Burden (RDB) for each patient. While some hubs exhibited uniform uncoupling across the cohort, others demonstrated pronounced subtype-specific rewiring, where the degree of uncoupling was characteristic of specific Integrative Clusters (*INTCLUST*). Ranking hubs by the interquartile range (IQR) of their median residuals across clusters revealed that the most highly rewired hubs included *SAPCD2* (IQR = 1.002), *RACGAP1* (IQR = 0.948), and *CDKN3* (IQR = 0.942). Notably, this rewired subset was highly enriched for cell-cycle regulators, including *MELK*, *PKMYT1*, and *BUB1*, suggesting that regulatory uncoupling at these nodes may facilitate the bypass of normal cellular checkpoints. Principal Component Analysis (PCA) on the top 50 rewired hubs yielded three orthogonal Discordance Factors (DF1, DF2, and DF3) for downstream evaluation.

3.3 Cohort reproducibility and structural stability

To ensure that the principal axes of discordance represent stable biological structures rather than cohort-specific artifacts, we evaluated their reproducibility across independent molecular accrual partitions (Cohort 1, $n = 387$; Cohort 3, $n = 495$). The primary axes of discordance demonstrated high structural stability (Figure 1A). The Pearson correlation between the independently derived gene loadings was 0.941 for the matched first principal component (PC1) and -0.917 for the matched second principal component (PC2). The third component exhibited lower stability (Pearson $R = 0.118$), yielding a mean absolute correlation of 0.6588 across the three matched components. Core genes that consistently defined the primary axis of discordance (PC1) in both cohorts included *CCNB2*, *UHRF1*, and *NUSAP1*. The high reproducibility of PC1 and PC2 confirms that the most dominant axes of multi-omic regulatory breakdown are robust, intrinsic features of breast cancer biology.

3.4 Mutational landscape of discordance

To determine whether the extracted discordance factors represent distinct biological processes or merely reflect general genomic instability, we evaluated their relationship with the mutational landscape. Mutation enrichment analysis revealed that the discordance state is strongly polarized by key driver mutations (Figure 1B). Specifically, DF2 was significantly elevated in tumors harboring *TP53* mutations (mean DF2: 0.788 in mutant vs. -0.417 in wild-type, $p = 2.04 \times 10^{-26}$). Conversely, *GATA3* mutations were associated with significantly lower DF2 scores (mean DF2: -0.773 in mutant vs. 0.118 in wild-type, $p = 1.42 \times 10^{-8}$). Despite these strong associations with specific drivers, the discordance factors exhibited only weak correlations with total mutational burden (Spearman ρ for DF1: 0.159, DF2: -0.089, DF3: -0.140). This confirms that multi-omic regulatory uncoupling is not simply a proxy for the overall quantity of mutations, but rather a distinct feature modulated by specific oncogenic pathways.

3.5 Cause-specific survival and lethality specificity

We tested the central hypothesis that localized multi-omic discordance specifically drives tumor lethality, rather than general patient frailty, using cause-specific Cox proportional hazards models. These models were rigorously adjusted for age, global copy-number burden, total mutational burden, tumor size, lymph node status, and existing taxonomies. The addition of DF1–DF3 to the clinical baseline model significantly improved the prediction of death from disease (Likelihood Ratio Test [LRT] statistic = 16.45, $p = 0.0009$), increasing the Concordance index (C-index) from 0.7196 to 0.7284. In contrast, the discordance factors failed to improve the prediction of death from other causes (LRT statistic = 2.86, $p = 0.4133$).

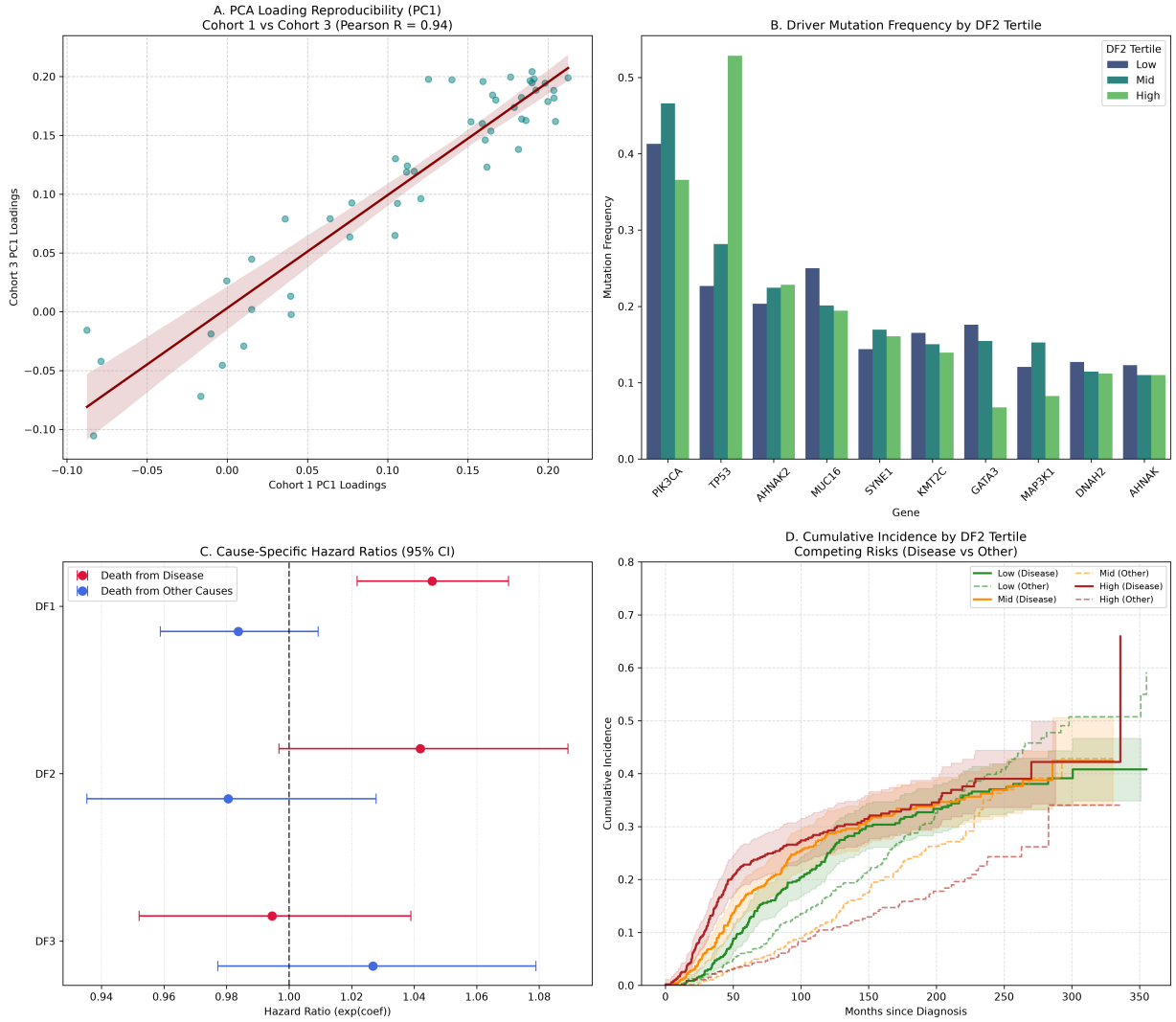


Figure 1: Multi-panel synthesis of multi-omic discordance factors (DF) and their role in breast cancer lethality. (A) PCA loading reproducibility for the first principal component (PC1) between Cohort 1 and Cohort 3 demonstrates high structural stability (Pearson R = 0.941), indicating that the primary axis of discordance is a reproducible biological feature rather than a cohort-specific artifact. (B) Driver mutation frequencies stratified by DF2 tertiles reveal that high DF2 is strongly associated with *TP53* mutations, whereas *GATA3* mutations are enriched in the low DF2 tertile, polarizing the discordance state. (C) Cause-specific hazard ratios (95% CI) demonstrate that DF1 significantly predicts death from disease but not death from other causes, highlighting lethality specificity. (D) Aalen-Johansen cumulative incidence functions stratified by DF2 tertiles confirm this specificity; the high DF2 tertile exhibits significantly elevated disease-specific mortality, while the incidence of death from other causes remains closely overlapping across all tertiles. Together, these results establish that the breakdown of transcriptional control is a specific, reproducible hallmark of lethal tumor biology rather than a proxy for general frailty.

Evaluating the individual factors revealed distinct prognostic roles (Figure 1C). DF1 emerged as a robust, independent predictor of disease-specific death (Hazard Ratio [HR] = 1.046 per unit increase, 95% CI: 1.022–1.070, $p = 0.00016$). DF2 demonstrated a trend toward increased disease-specific hazard (HR = 1.042, 95% CI: 0.997–1.089, $p = 0.069$) but showed absolutely no association with other-cause mortality (HR = 0.981, $p = 0.413$).

This lethality specificity was further confirmed through Aalen-Johansen Cumulative Incidence Function (CIF) analysis under competing risks (Figure 1D). Stratification of the cohort by DF2 tertiles revealed a marked separation in the cumulative incidence of disease-specific death, with the high-discordance tertile experiencing significantly elevated mortality. Crucially, the CIF curves for death from other causes remained closely overlapping across all tertiles, definitively establishing high multi-omic discordance as a specific marker of cancer lethality rather than a reflection of general comorbidities.

4 Conclusions

In this study, we addressed a critical limitation in breast cancer genomics: the tendency for macroscopic genomic alterations to obscure localized, multi-omic regulatory failures. By systematically identifying "discordant hubs"—genes that have lost their baseline transcriptional coupling to DNA copy number and promoter methylation—we isolated a distinct biological axis of tumor aggressiveness. Our findings demonstrate that this localized regulatory uncoupling is not merely a byproduct of global genomic instability, but rather a structured and reproducible hallmark of lethal tumor biology.

The biological relevance of this regulatory breakdown is underscored by its subtype-specific rewiring and its strong association with key oncogenic drivers. The most highly rewired discordant hubs are prominently enriched for cell-cycle regulators, suggesting that the uncoupling of these specific nodes from their genomic and epigenetic scaffolds may facilitate the bypass of normal cellular checkpoints. Furthermore, the discordance state is heavily polarized by mutations in critical genes such as *TP53* and *GATA3*, yet it remains independent of the total mutational burden. This indicates that multi-omic discordance reflects specific, pathway-driven transcriptional dysregulation rather than general genomic decay.

Crucially, our competing risks framework establishes the clinical utility and specificity of the Regulatory Discordance Burden. By explicitly separating disease-specific mortality from other-cause mortality, we confirmed that the primary axes of discordance (particularly DF1) significantly improve the prediction of cancer lethality without proxying general patient frailty or aging. The addition of these factors provided orthogonal prognostic value beyond established clinical covariates, global copy-number burden, and existing transcriptomic taxonomies.

Finally, the high structural reproducibility of the primary discordance axes across independent patient cohorts confirms that these localized regulatory failures are stable, intrinsic features of aggressive breast cancer. Ultimately, the identification and quantification of multi-omic discordant hubs provide a refined, cause-specific predictor of tumor lethality. These insights not only improve prognostic modeling but also highlight the breakdown of transcriptional control as a critical feature of lethal breast cancer, paving the way for future investigations into the mechanistic consequences of uncoupled regulatory nodes.