Tumor-Specific Buffering State Vectors Derived from Dosage-Sensitive Hubs as Intra-Cluster Predictors of Disease-Specific Mortality

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Breast cancer exhibits heterogeneous transcriptional rewiring in response to genomic copy number alterations, yet quantifying this plasticity without relying on static pathway aggregations remains challenging. To address this, we introduce a framework that quantifies tumor-specific transcriptional buffering by evaluating deviations from expected dosage-expression relationships. Using matched genomic and transcriptomic data from 1,980 METABRIC breast cancer samples, we identified 500 dosage-sensitive hubs and applied principal component analysis to robust Huber regression residuals to construct multidimensional buffering state vectors for each tumor. We then evaluated the prognostic utility of these vectors for predicting disease-specific mortality using Fine-Gray competing risks survival models, adjusting for age at diagnosis strictly within established integrative clusters. Our analysis revealed that these state vectors capture variance orthogonal to static classifications and significantly refine prognosis within the 4ER+ and 8 integrative clusters. Notably, a distinct buffering state in the 4ER+ subtype demonstrated a strong protective effect against disease-specific death that successfully replicated in an independent validation cohort split, establishing the transcriptional buffering of dosage-sensitive hubs as a reproducible, intra-cluster axis of lethality.

1 Introduction

Genomic copy number alterations (CNAs) are fundamental drivers of breast cancer heterogeneity, profoundly shaping the transcriptomic landscape and influencing tumor evolution. However, the relationship between gene dosage and mRNA expression is rarely a simple linear translation. Tumors frequently exhibit transcriptional plasticity, rewiring their regulatory networks to either buffer against or exacerbate the effects of genomic amplifications and deletions. Traditional approaches to stratifying breast cancer often rely on static pathway aggregations or global copy-number burden normalizations. While these methods have successfully established robust taxonomies, such as integrative clusters, they inherently obscure the dynamic, tumor-specific transcriptional responses to genomic dosage.

Quantifying this transcriptional plasticity remains a significant computational challenge. Global normalizations tend to strip away subtype-intrinsic biological signals, while static gene sets fail to capture the heterogeneous ways in which individual tumors modulate their genomic burden. To fully understand the molecular pathology of breast cancer, it is necessary to evaluate deviations from expected dosage-expression relationships at the level of individual tumors, thereby isolating the active rewiring processes from the static genomic background.

To address this gap, we introduce a novel analytical framework that quantifies tumor-specific transcriptional buffering. We hypothesize that certain genes act as "dosage-sensitive hubs"—loci whose expression is tightly coupled to their copy number state across the broader patient population. By modeling the expected dosage-expression relationship for these hubs using robust regression, we can extract the residuals for each tumor. These residuals represent the degree of transcriptional buffering or exacerbation. Applying dimensionality reduction to these residuals yields multidimensional "buffering state vectors," which serve as a quantitative signature of a tumor's unique transcriptional response to its genomic dosage.

In this study, we apply this framework to matched genomic and transcriptomic data from 1,980 breast cancer samples. We identify 500 dosage-sensitive hubs and construct buffering state vectors for each sample. To rigorously test whether these vectors capture clinically relevant variance orthogonal to existing static classifications, we evaluate their prognostic utility strictly within established integrative clusters. We employ Fine-Gray competing risks survival models to explicitly isolate disease-specific mortality from other causes of death, adjusting for age at diagnosis. Furthermore, we assess the structural stability and prognostic independence of these vectors across distinct accrual cohort partitions to ensure reproducibility.

Our analysis demonstrates that buffering state vectors capture significant biological variance missed by static taxonomies. We show that these vectors significantly refine prognosis within specific subtypes, most notably the 4ER+ and 8 integrative clusters. Crucially, we identify a distinct buffering state within the 4ER+ subtype that confers a strong protective effect against disease-specific death, a finding that successfully replicates in an independent validation cohort split. Ultimately, this work establishes the transcriptional buffering of dosage-sensitive hubs as a reproducible, intra-cluster axis of lethality, providing a new lens through which to view the joint organization of molecular layers in breast cancer.

2 Methods

2.1 Data alignment and preprocessing

Matched clinical, transcriptomic, and genomic copy number alteration (CNA) data were obtained for the METABRIC cohort. Expression data, provided as log2-scaled microarray intensities, were utilized without additional count-based normalizations to preserve intrinsic variance. The dataset was strictly subset to 1,980 samples possessing complete and aligned records across the clinical, expression, and CNA layers.

To focus the analysis on genes with robust biological signals, the intersection of genes present in both the expression and CNA matrices was determined. The per-gene expression variance was computed across all 1,980 samples, and the dataset was filtered to retain the top 5,000 most variable genes. A secondary filter was applied to ensure that these 5,000 genes exhibited a non-zero CNA value in at least 5% of the samples, thereby removing genes with negligible genomic variation.

For survival analysis, a competing risks target variable was constructed from the vital status records. Patients were categorized as living (0), died of disease (1, the primary event of interest), or died of other causes (2, the competing event). Overall survival in months was extracted as the time-to-event variable.

2.2 Identification of dosage-sensitive hubs

To identify genes whose transcription is most strictly driven by genomic dosage across the cohort, the Pearson correlation coefficient between the copy number state and the expression level was computed for each of the 5,000 filtered genes across all 1,980 samples. The distribution of these correlation coefficients was evaluated, and "dosage-sensitive hubs" (DSHs) were defined as the genes falling into the top 10% of positive CNA-expression correlations. This procedure isolated 500 genes that serve as robust sensors for quantifying transcriptional buffering.

2.3 Construction of tumor-specific buffering state vectors

To quantify how individual tumors deviate from the expected dosage response without introducing biases from global copy-number burden normalizations, a robust regression approach was employed. For each of the 500 DSHs, a Huber regression model was fitted to predict expression from CNA across all samples. The residuals from these models were extracted, representing the tumor-specific buffering (or exacerbation) of the expected dosage effect for each gene.

These residuals were assembled into a matrix of 1,980 samples by 500 hubs. Principal Component Analysis (PCA) was subsequently applied to this residual matrix to reduce dimensionality. The top principal components (PCs) that cumulatively explained at least 80% of the variance were extracted. These components constitute the multidimensional "buffering state vector" for each tumor. To aid biological interpretability, the top components were correlated with the original gene residuals to identify the specific genes driving each vector. Furthermore, analysis of variance (ANOVA) was used to assess the association between the

top components and established integrative cluster (*INTCLUST*) labels, verifying that the buffering state vectors capture orthogonal information rather than merely re-encoding static classifications.

2.4 Intra-cluster competing risks survival modeling

The prognostic utility of the buffering state vectors was evaluated strictly within the precomputed integrative clusters to prevent circularity and demonstrate intra-cluster resolution. The state vectors were merged with the processed clinical data, and the analysis was iterated across the 11 *INTCLUST* groups. Clusters with fewer than 50 samples or those where models failed to converge were pooled with clinically similar clusters (e.g., 4ER+ and 4ER-) or excluded if insufficient data remained.

For each valid cluster, a Fine-Gray competing risks survival model was fitted. Disease-specific mortality was treated as the primary event, and death from other causes was treated as a competing risk, utilizing overall survival months as the time scale. The cluster-specific models included the buffering state vector components and age at diagnosis as covariates. To formally test the added prognostic value of the state vectors, a Likelihood Ratio Test (LRT) was performed within each cluster, comparing a baseline model containing only age at diagnosis against the full model containing both age and the state vectors. Subdistribution hazard ratios (SHRs), 95% confidence intervals, and p-values were extracted for the state vector components.

2.5 Cross-cohort replication and stability assessment

To validate the structural stability and prognostic independence of the buffering state vectors, the 1,980 samples were partitioned along the accrual axis using the cohort variable. The dataset was split into a discovery macro-cohort (Cohorts 1–3) and a validation macro-cohort (Cohorts 4–9), ensuring sufficient sample sizes within the *INTCLUST* groups for both partitions.

The intra-cluster Fine-Gray competing risks models were first fitted in the discovery set to identify significant prognostic components. Subsequently, these models were refitted independently within the validation set to test the specific components identified as significant in the discovery phase. Replication was assessed by comparing the SHRs of the buffering state vector components between the discovery and validation partitions. A true biological signal was required to maintain its direction of effect across the accrual axis and achieve statistical significance ($p < 0.05$) in the validation set.

3 Results

3.1 Identification of dosage-sensitive hubs

The primary objective of our framework was to define a set of dosage-sensitive hubs (DSHs) characterized by the highest degree of transcriptional coupling with genomic copy number alterations (CNA). By computing Pearson correlation coefficients between copy number states and expression levels across the 1,980 aligned METABRIC samples for the 5,000 most variable genes (with a secondary filter for at least 5% non-zero CNA values), we evaluated the baseline dosage-expression relationship. The distribution of these correlation coefficients revealed a median coupling of 0.117 across the filtered gene set (Figure 1).

By isolating the top 10% of positive correlations, we identified 500 DSHs. This top-ranked hub set is highly enriched for well-characterized oncogenic drivers and chromosome 17q amplicons, including `PGAP3` ($r = 0.779$), `LSM1` ($r = 0.776$), `ASH2L` ($r = 0.767$), and `ERBB2` ($r = 0.765$), alongside other prominent hubs such as `FADD` ($r = 0.764$), `GRB7` ($r = 0.763$), and `STARD3` ($r = 0.715$). These high correlations confirm that the DSH set comprises genes whose mRNA abundance is strictly sensitive to DNA dosage, providing an ideal foundation for quantifying tumor-specific transcriptional buffering.

3.2 Variance capture and orthogonality of buffering state vectors

To quantify tumor-specific buffering, we constructed a residual matrix by fitting robust Huber regressions for each DSH and extracting the residuals, which represent the observed expression minus the dosage-predicted expression. Principal Component Analysis (PCA) was applied to this residual matrix to derive

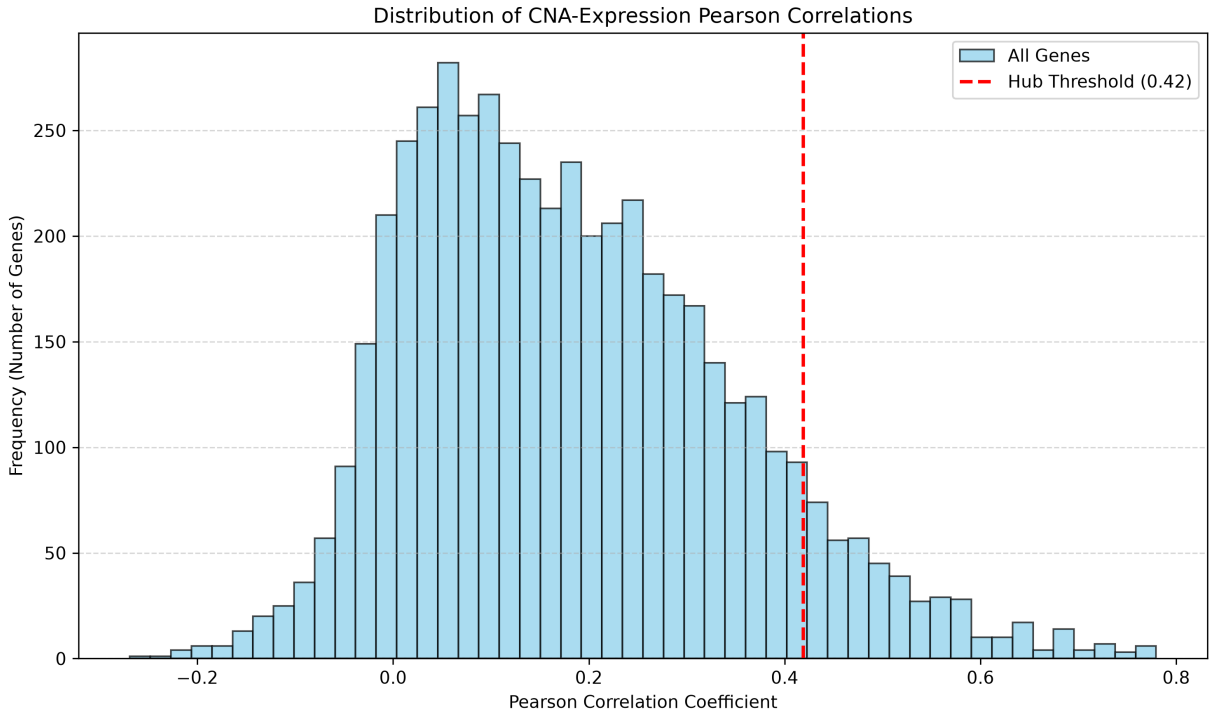


Figure 1: Distribution of Pearson correlation coefficients between genomic copy number alterations (CNA) and mRNA expression levels for the 5,000 most variable genes. The histogram illustrates a baseline median coupling of 0.117 across the filtered gene set. The red dashed line at $r = 0.42$ marks the threshold used to isolate the top 10% (500 genes) as Dosage-Sensitive Hubs (DSHs). The presence of a distinct right tail reveals a subset of genes, including known oncogenic drivers like `ERBB2` ($r = 0.765$), whose abundance is strictly sensitive to DNA dosage. Isolating these highly coupled genes demonstrates that while most genes exhibit weak dosage dependency, this specific subset provides the necessary sensors to quantify tumor-specific transcriptional buffering and deviation from expected dosage-expression relationships.

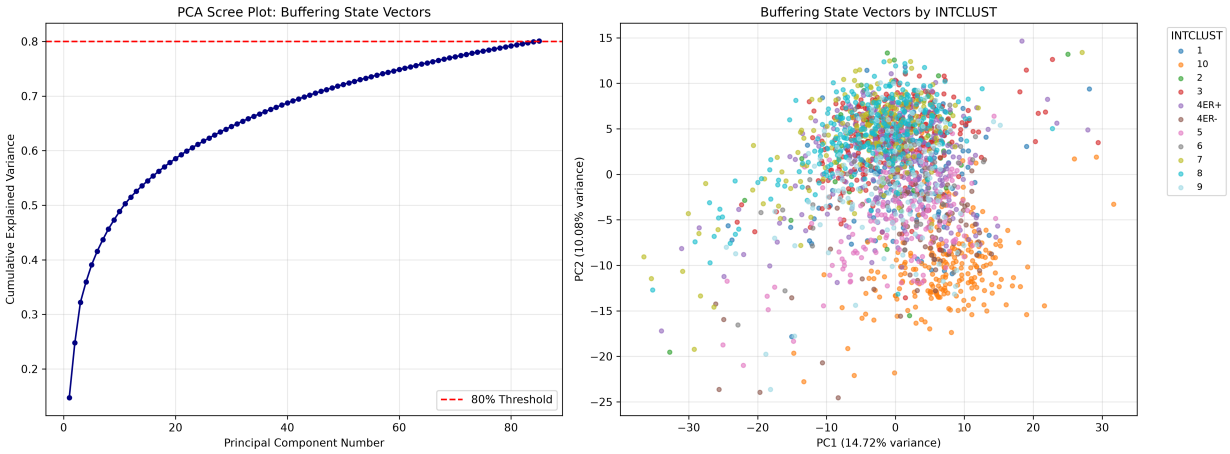


Figure 2: Principal Component Analysis (PCA) of tumor-specific buffering state vectors. The left panel displays the cumulative explained variance across principal components, with a red dashed line indicating the 80% threshold. The right panel presents a scatter plot of the first two principal components, colored by `INTCLUST` subtype. Tumors within the same integrative cluster occupy a broad range of the buffering state space, demonstrating that these vectors capture heterogeneous biological states and provide an intra-cluster resolution of transcriptional rewiring that is partially orthogonal to existing static classifications.

multidimensional buffering state vectors. The top five principal components (PCs) collectively captured a substantial portion of the variance in dosage-expression discordance (Figure 2).

Analysis of the PCA loadings revealed distinct biological orientations; for instance, `Buffering_PC1` is heavily influenced by genes such as `MIEN1` (loading -0.428) and `ERBB2` (loading -0.354), indicating that it captures buffering states related to the 17q12-21 amplicon. Crucially, these vectors represent variance that is partially orthogonal to existing taxonomies. ANOVA-based associations with `INTCLUST` labels indicated that while some clustering by subtype exists, the state vectors provide an intra-cluster resolution of transcriptional rewiring. Scatter plots of the first two components colored by `INTCLUST` demonstrate that tumors within the same integrative cluster occupy a broad range of the buffering state space. This confirms that the buffering state vectors capture heterogeneous biological states and intra-cluster transcriptional rewiring that static classifications miss.

3.3 Prognostic value for disease-specific mortality

We evaluated the prognostic utility of the buffering state vectors using Fine-Gray competing risks models. The cohort included 646 disease-specific mortality events (the primary target) and 496 deaths from other causes (the competing risk) (Figure 3). Models were fitted independently within `INTCLUST` groups, adjusting for `AGE_AT_DIAGNOSIS`.

Likelihood Ratio Tests (LRT) comparing a baseline age-only model to the full model (age plus the top five PCs) identified two clusters where buffering states significantly refined prognosis: `INTCLUST` 4ER+ (LRT $p = 4.61 \times 10^{-7}$, $N = 260$, 65 events) and `INTCLUST` 8 (LRT $p = 1.06 \times 10^{-2}$, $N = 299$, 82 events). Within `INTCLUST` 4ER+ (Figure 4), `Buffering_PC2` emerged as a strong protective factor against disease-specific death with a subdistribution hazard ratio (SHR) of 0.846 (95% CI: 0.798–0.897, $p = 2.20 \times 10^{-8}$), while `Buffering_PC5` was associated with increased risk (SHR = 1.129, 95% CI: 1.050–1.215, $p = 1.11 \times 10^{-3}$).

Similarly, in `INTCLUST` 8 (Figure 5), `Buffering_PC2` demonstrated a protective effect (SHR = 0.889, 95% CI: 0.833–0.948, $p = 3.71 \times 10^{-4}$), whereas `Buffering_PC1` was associated with higher risk (SHR = 1.038, 95% CI: 1.002–1.075, $p = 0.036$).

The prognostic impact of these buffering states was highly subtype-specific. While `Buffering_PC2` showed nominally significant protective trends in several other clusters, including `INTCLUST` 1 (Figure 6), `INTCLUST` 3 (Figure 7), and `INTCLUST` 6 (Figure 8), the overall multi-component models did not reach global significance in these groups.

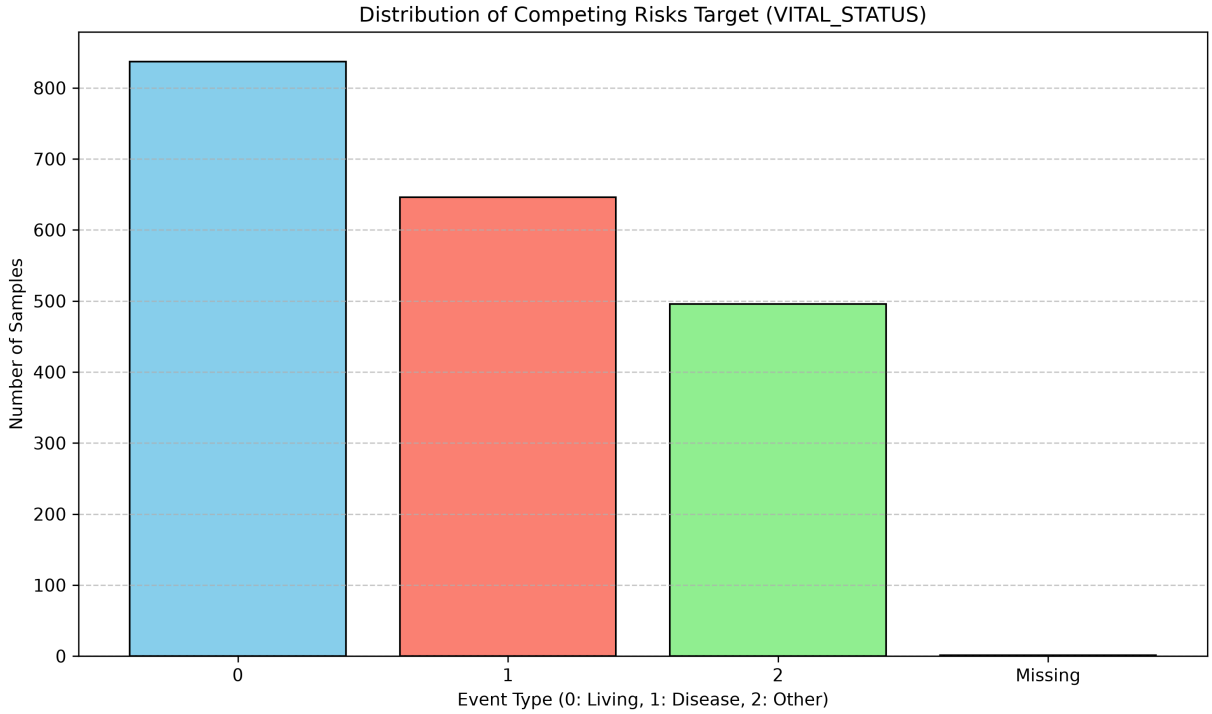


Figure 3: Distribution of the competing risks target variable (`VITAL_STATUS`) across the 1,980 METABRIC samples. The bar chart illustrates the frequency of patient outcomes categorized as living (0), died of disease (1), died of other causes (2), and missing. As utilized in the Fine-Gray competing risks models to evaluate the prognostic value of the Buffering State Vectors, the cohort includes 646 disease-specific mortality events (the primary target) and 496 deaths from other causes (the competing risk). This distribution highlights the substantial proportion of competing mortality events that must be accounted for to accurately assess the impact of tumor-specific transcriptional buffering states on disease-specific survival without confounding from other causes of death.

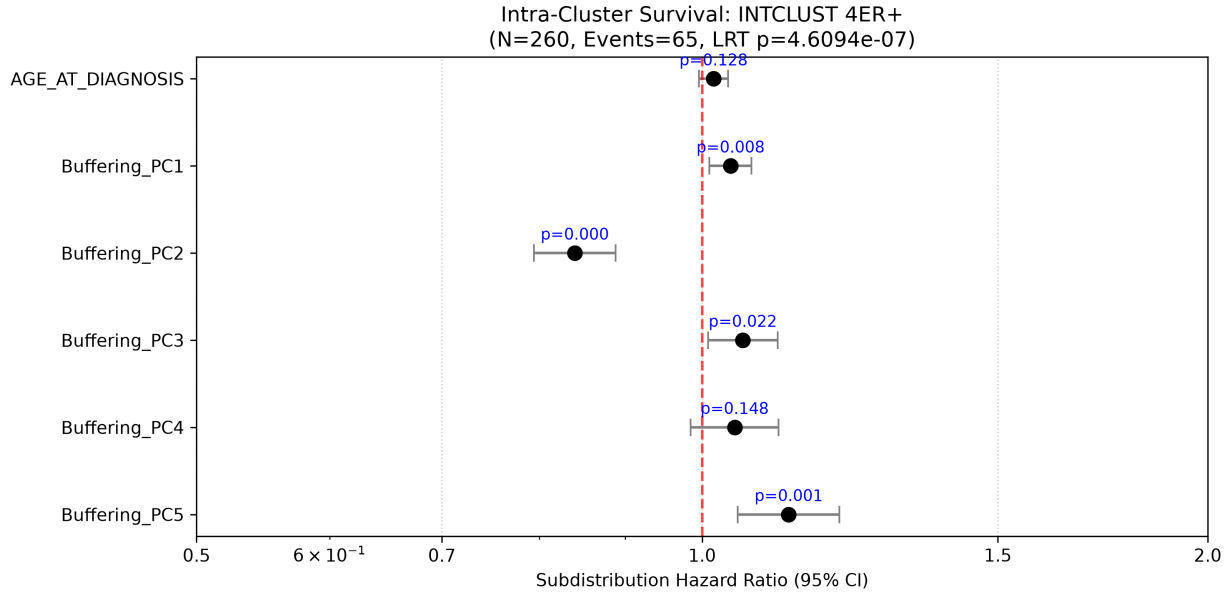


Figure 4: Forest plot displaying Subdistribution Hazard Ratios (95% CI) from a Fine-Gray competing risks model evaluating the prognostic impact of Buffering State Vectors on disease-specific mortality within the `INTCLUST` 4ER+ subtype ($N = 260$, 65 events). The model adjusts for `AGE_AT_DIAGNOSIS` alongside the top five principal components (`Buffering_PC1` through `Buffering_PC5`) representing dosage-expression discordance. The inclusion of these buffering states significantly refines prognosis compared to an age-only baseline (LRT $p = 4.61 \times 10^{-7}$). The plot reveals that `Buffering_PC2` acts as a strong protective factor, indicating effective buffering where tumors fail to translate genomic gains into transcriptional advantages. In contrast, `Buffering_PC5` is significantly associated with increased mortality risk, likely reflecting a breakdown in dosage-control mechanisms that exacerbates oncogenic potential. These findings demonstrate that intra-cluster buffering states capture critical, heterogeneous transcriptional rewiring that static genomic classifications miss.

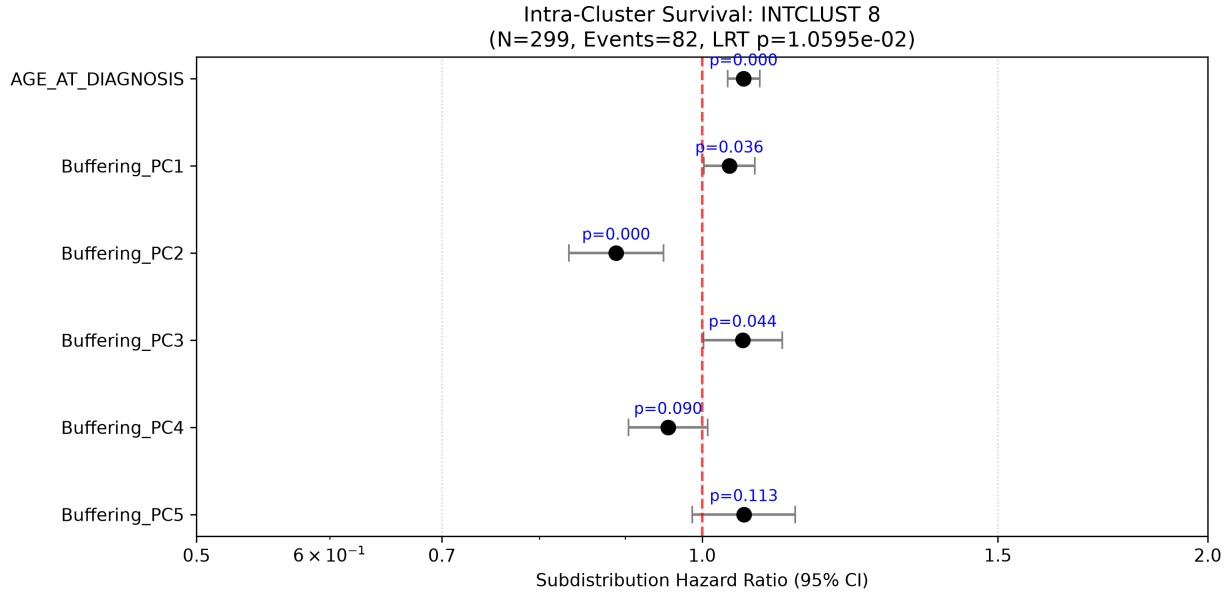


Figure 5: Forest plot displaying Subdistribution Hazard Ratios (SHR) and 95% confidence intervals from a Fine-Gray competing risks model evaluating disease-specific mortality within the `INTCLUST` 8 subtype ($N = 299$, 82 events). The model assesses the prognostic value of the top five Buffering State Vectors, which quantify tumor-specific deviations from expected dosage-expression relationships, adjusted for `AGE_AT_DIAGNOSIS`. The inclusion of these vectors significantly improves prognostic accuracy over an age-only baseline (Likelihood Ratio Test $p = 1.06 \times 10^{-2}$). We observe that `Buffering_PC2` confers a significant protective effect (SHR $= 0.889$, $p = 3.71 \times 10^{-4}$), likely reflecting effective buffering where the tumor fails to translate genomic gains into a transcriptional advantage, whereas `Buffering_PC1` is associated with increased risk (SHR $= 1.038$, $p = 0.036$). This demonstrates that intra-cluster transcriptional plasticity and the rewiring of dosage-sensitive hubs provide critical, independent axes of lethality not captured by static copy-number taxonomies.

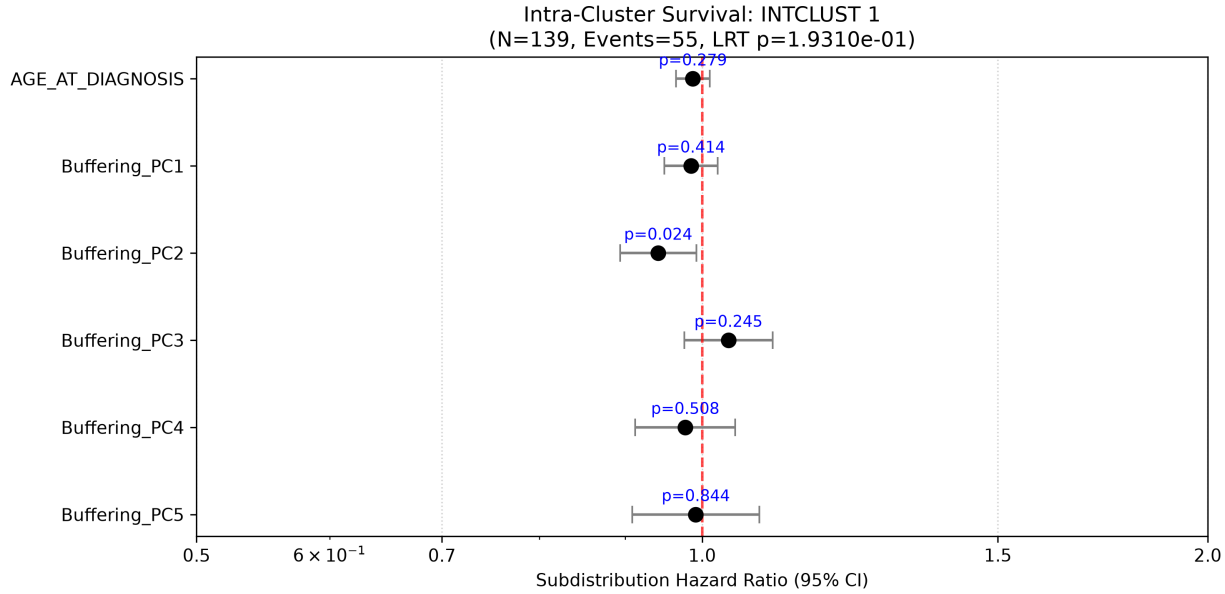


Figure 6: Forest plot displaying the intra-cluster prognostic evaluation of Buffering State Vectors within the `INTCLUST` 1 subtype ($N = 139$, 55 events) using a Fine-Gray competing risks model. The analysis evaluates the Subdistribution Hazard Ratios (SHR) and 95% confidence intervals for the first five principal components (`Buffering_PC1` through `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. Unlike other clusters where buffering states significantly refined prognosis, the overall likelihood ratio test (LRT) comparing this full model to an age-only baseline is not statistically significant (LRT $p = 0.193$). However, `Buffering_PC2` demonstrates a nominally significant protective effect against disease-specific mortality ($p = 0.024$, SHR < 1). This suggests that while the collective buffering state vectors do not broadly predict outcomes within `INTCLUST` 1, specific axes of transcriptional rewiring—potentially representing effective buffering where tumors fail to translate genomic alterations into transcriptional advantages—may still capture underlying biological heterogeneity.

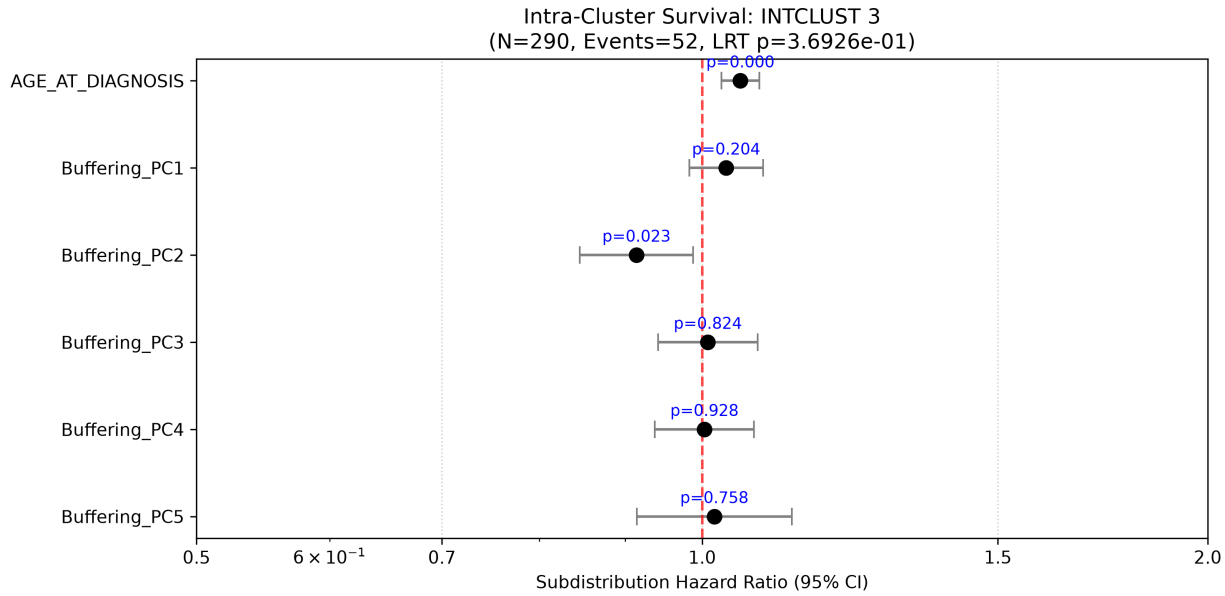


Figure 7: Forest plot displaying the subdistribution hazard ratios and 95% confidence intervals from a Fine-Gray competing risks model for disease-specific mortality within the `INTCLUST` 3 cohort ($N = 290$, 52 events). The model evaluates the prognostic utility of five Buffering State Vectors (`Buffering_PC1` to `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. Although the likelihood ratio test (LRT) comparing this full model to an age-only baseline is not statistically significant (LRT $p = 0.369$), `Buffering_PC2` demonstrates a significant protective effect ($p = 0.023$). This indicates that specific transcriptional buffering states can capture intra-cluster survival heterogeneity and modulate the oncogenic potential of genomic alterations, even when the aggregate multi-component model does not reach global significance.

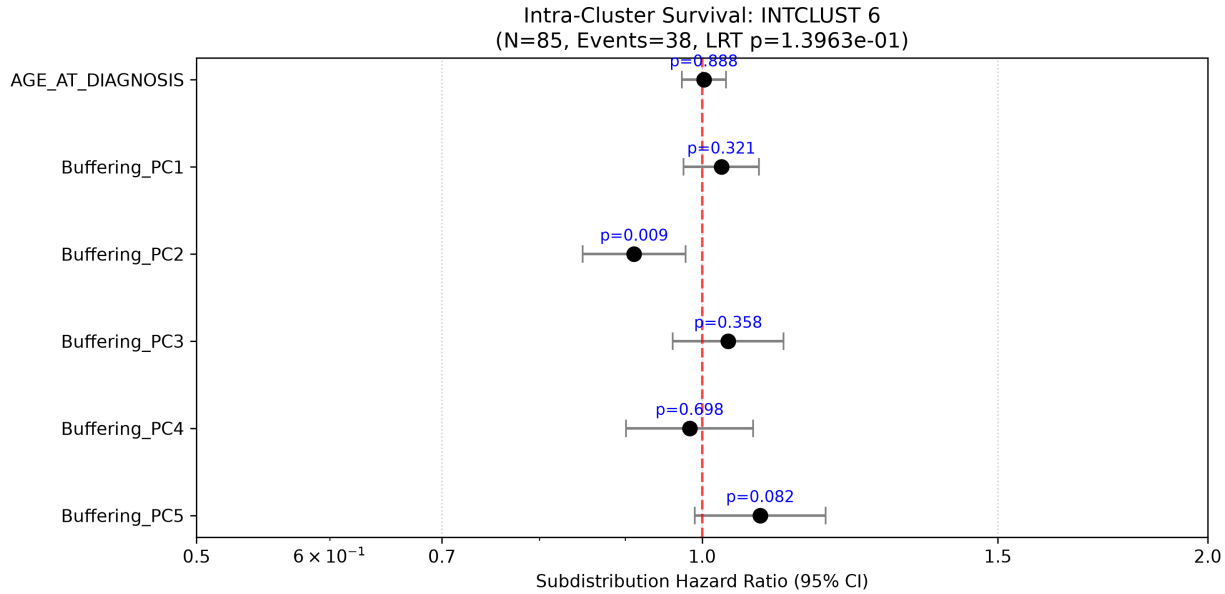


Figure 8: Forest plot displaying subdistribution hazard ratios (SHR) and 95% confidence intervals from a Fine-Gray competing risks model for disease-specific mortality within the `INTCLUST` 6 subtype ($N = 85$, 38 events). The model evaluates five Buffering State Vectors (`Buffering_PC1` through `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. Although the overall likelihood ratio test (LRT $p = 0.140$) shows that the full model does not significantly improve upon an age-only baseline for this specific cluster, `Buffering_PC2` demonstrates a significant protective effect ($p = 0.009$, SHR < 1). This finding aligns with the protective role of `Buffering_PC2` identified in other integrative clusters, indicating that this specific transcriptional state—where a tumor likely fails to translate genomic gains into an oncogenic advantage—may confer a survival benefit across multiple breast cancer subtypes despite heterogeneous overall cluster significance.

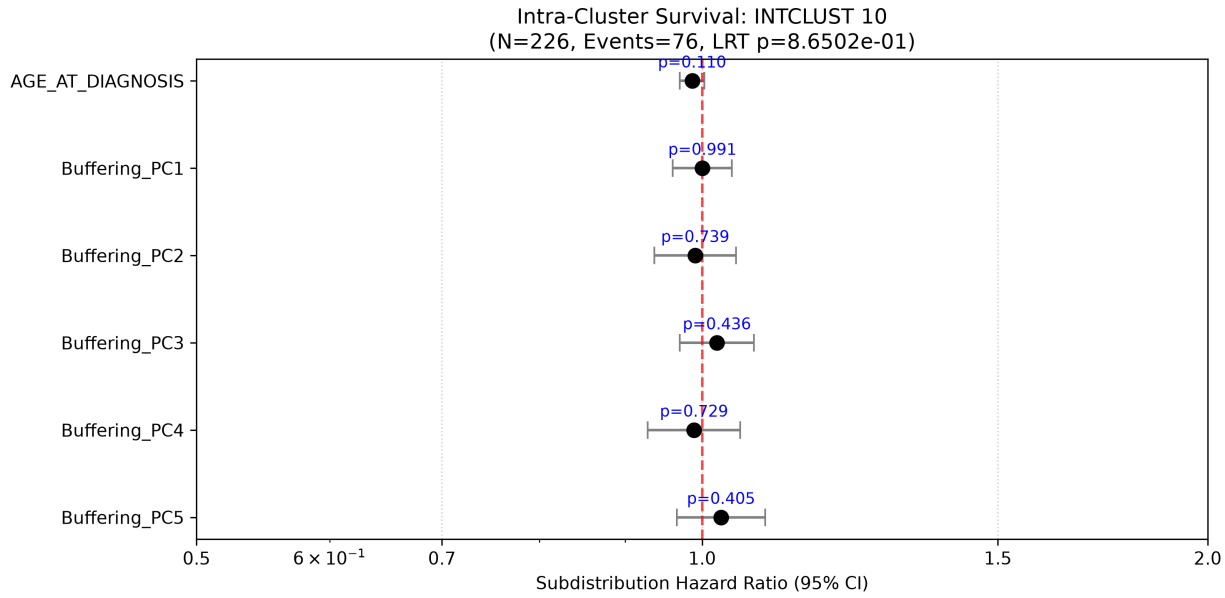


Figure 9: Forest plot detailing the results of a Fine-Gray competing risks model for disease-specific mortality within the INTCLUST 10 subtype ($N = 226$, 76 events). The plot displays the subdistribution hazard ratios and 95% confidence intervals for `AGE_AT_DIAGNOSIS` and the top five Buffering State Vectors (`Buffering_PC1` to `Buffering_PC5`), which quantify tumor-specific transcriptional buffering of dosage-sensitive hubs. In contrast to the significant prognostic value observed in other integrative clusters, the overall Likelihood Ratio Test (LRT $p = 0.865$) and individual feature p-values demonstrate that these buffering states do not significantly refine prognosis within INTCLUST 10. This indicates that the impact of dosage-expression discordance on patient outcomes is heterogeneous and specific to certain genomic subtypes.

Furthermore, buffering state vectors did not significantly improve upon the baseline age-only model in `INTCLUST` 10 (Figure 9), `INTCLUST` 2 (Figure 10), `INTCLUST` 4ER- (Figure 11), `INTCLUST` 5 (Figure 12), `INTCLUST` 7 (Figure 13), or `INTCLUST` 9 (Figure 14). This heterogeneity underscores that the lethal consequences of dosage-expression discordance are deeply context-dependent.

3.4 Cross-cohort replication and stability

To assess the stability of these findings across the accrual axis, we partitioned the cohort into a discovery set (Cohorts 1–3, $N = 1,571$) and a validation set (Cohorts 4–9, $N = 408$). Replication was strictly defined as maintaining a significant ($p < 0.05$) association in the same direction within the validation set for features that were significant in the discovery set. Across the eight buffering components identified as significant in the discovery set across all clusters, five (62.5%) maintained directional consistency in the validation set.

Notably, `Buffering_PC2` within `INTCLUST` 4ER+ was the only feature to achieve strict statistical significance in the validation split (Figure 15). In the discovery set, its SHR was 0.846 ($p = 2.20 \times 10^{-8}$); in the validation set, the protective effect was even more pronounced (SHR = 0.632, 95% CI: 0.416–0.961, $p = 0.032$).

Other signals exhibited directional consistency but failed to reach significance, likely constrained by the smaller sample size and lower event rates of the validation partition ($N = 408$ vs. $N = 1,571$). For instance, `Buffering_PC2` in `INTCLUST` 8 (Figure 16) yielded a discovery SHR of 0.899 ($p = 2.75 \times 10^{-3}$) and a validation SHR of 0.773 ($p = 0.053$), narrowly missing the significance threshold. Similarly, `Buffering_PC1` in `INTCLUST` 4ER+ showed a discovery SHR of 1.036 ($p = 0.018$) and a validation SHR of 1.439 ($p = 0.161$).

Cross-cohort replication analyses for the remaining clusters further highlighted the challenges of validating intra-cluster transcriptional buffering effects in smaller partitions, with none of the buffering components achieving statistical significance in the validation sets for `INTCLUST` 1 (Figure 17), `INTCLUST` 10 (Figure

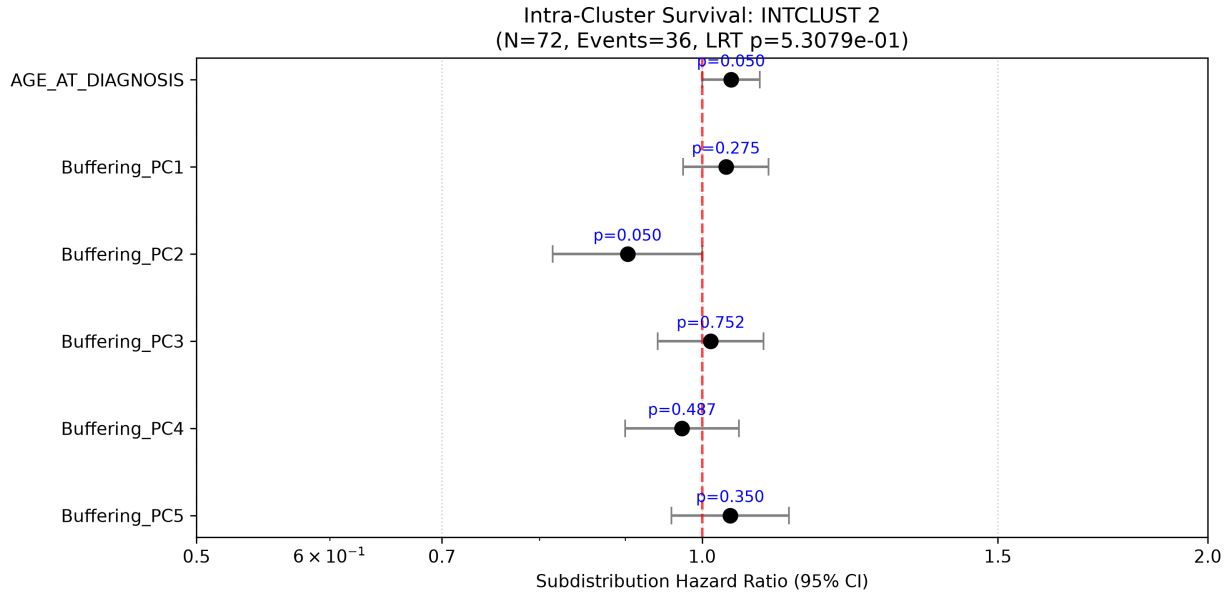


Figure 10: Forest plot displaying the Subdistribution Hazard Ratios (95% CI) from a Fine-Gray competing risks model for disease-specific mortality within the `INTCLUST` 2 subgroup ($N = 72$, 36 events). The model evaluates the prognostic value of the first five Buffering State Vectors (`Buffering_PC1` through `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. The overall Likelihood Ratio Test (LRT) yields $p = 0.53$, indicating that the inclusion of these buffering states does not significantly refine prognosis over age alone in this specific integrative cluster. Although `Buffering_PC2` exhibits a borderline protective trend ($p = 0.050$), the buffering of dosage-sensitive hubs does not provide robust intra-cluster prognostic value for `INTCLUST` 2, contrasting with the significant survival associations observed in other subtypes such as `INTCLUST` 4ER+ and `INTCLUST` 8.

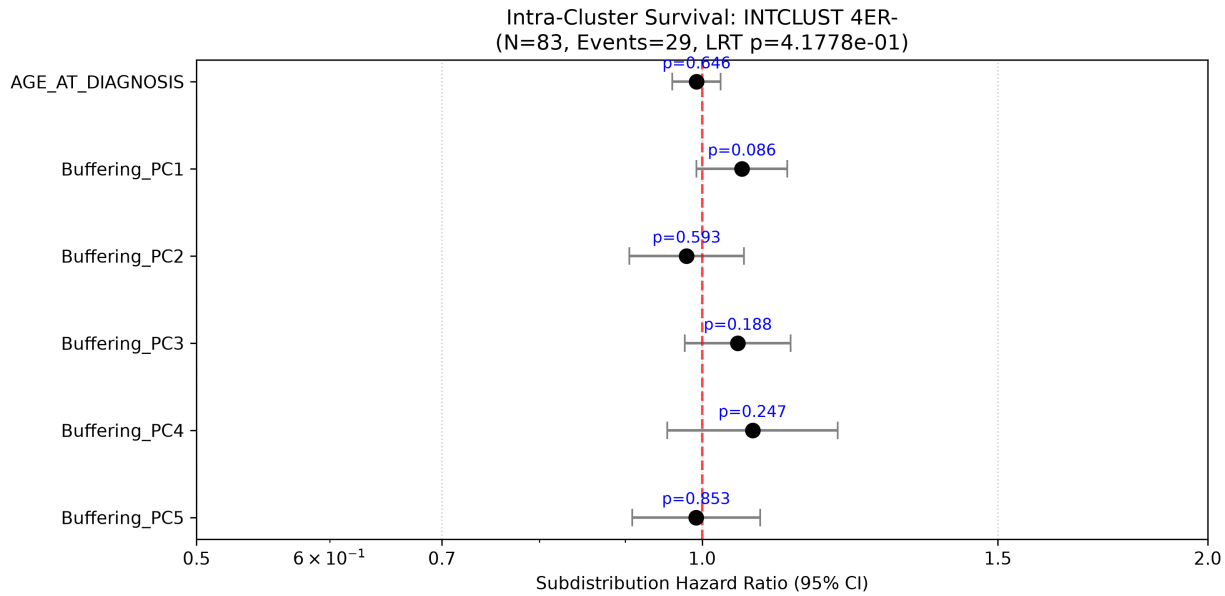


Figure 11: Forest plot displaying Subdistribution Hazard Ratios (95% CI) from a Fine-Gray competing risks model evaluating disease-specific mortality within the `INTCLUST` 4ER- subgroup ($N = 83$, 29 events). The model assesses the prognostic value of the top five Buffering State Vectors (`Buffering_PC1` through `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. Unlike the significant associations observed in other clusters, the overall Likelihood Ratio Test (LRT $p = 0.418$) and individual p-values demonstrate that these principal components do not significantly improve upon the baseline age-only model, as all confidence intervals cross the null effect line at 1.0. This indicates that while the buffering of dosage-sensitive hubs is a strong predictor of lethality in certain subtypes, it does not provide significant prognostic value for capturing transcriptional rewiring or disease progression within the 4ER- integrative cluster.

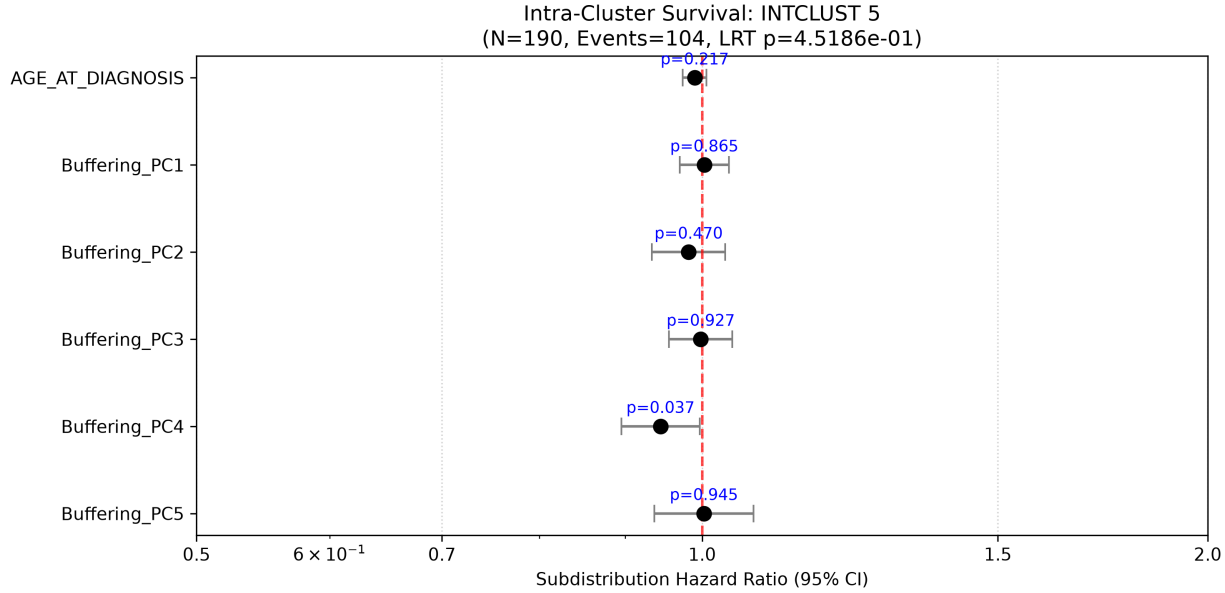


Figure 12: Forest plot of intra-cluster survival analysis for the `INTCLUST` 5 subtype ($N = 190$, 104 events) using a Fine-Gray competing risks model for disease-specific mortality. The model evaluates the prognostic utility of the top five Buffering State Vectors (`Buffering_PC1` to `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. Points represent the Subdistribution Hazard Ratio (SHR) with 95% confidence intervals, and individual p-values are annotated above each estimate. While the overall Likelihood Ratio Test (LRT) comparing this full model to an age-only baseline is not statistically significant ($p = 0.452$), `Buffering_PC4` shows a nominally significant protective effect ($p = 0.037$, SHR < 1). This indicates that although the collective buffering state vectors do not significantly refine prognosis across the entire `INTCLUST` 5 cohort as they do in other subtypes, specific axes of transcriptional rewiring and dosage-expression discordance can still capture localized heterogeneity in tumor lethality.

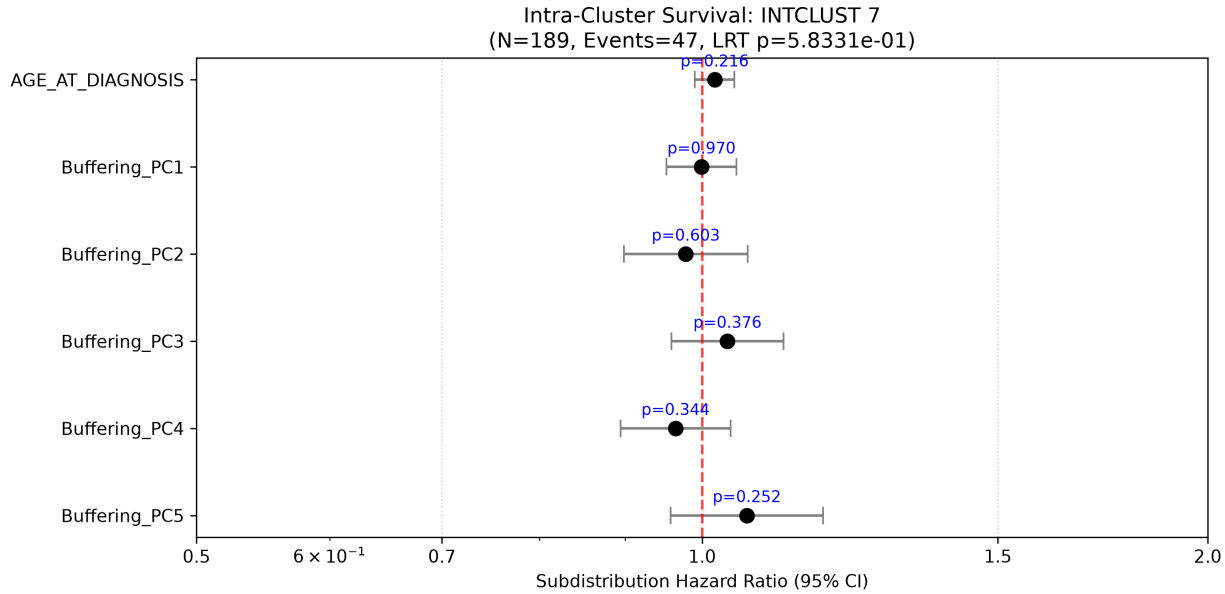


Figure 13: Forest plot detailing the intra-cluster survival analysis for `INTCLUST` 7 ($N = 189$, 47 events) using a Fine-Gray competing risks model for disease-specific mortality. The plot displays the subdistribution hazard ratios and 95% confidence intervals for `AGE_AT_DIAGNOSIS` and the top five Buffering State Vectors (`Buffering_PC1` to `Buffering_PC5`). Unlike the significant prognostic refinement observed in subtypes such as `INTCLUST` 4ER+ and 8, all confidence intervals for `INTCLUST` 7 cross the null effect line of 1.0, and individual component p-values are non-significant. Consequently, the overall Likelihood Ratio Test comparing this full model to an age-only baseline is not significant (LRT $p = 0.583$), indicating that these specific dosage-expression buffering states do not provide additional prognostic value or capture a lethal axis of transcriptional rewiring within this particular integrative cluster.

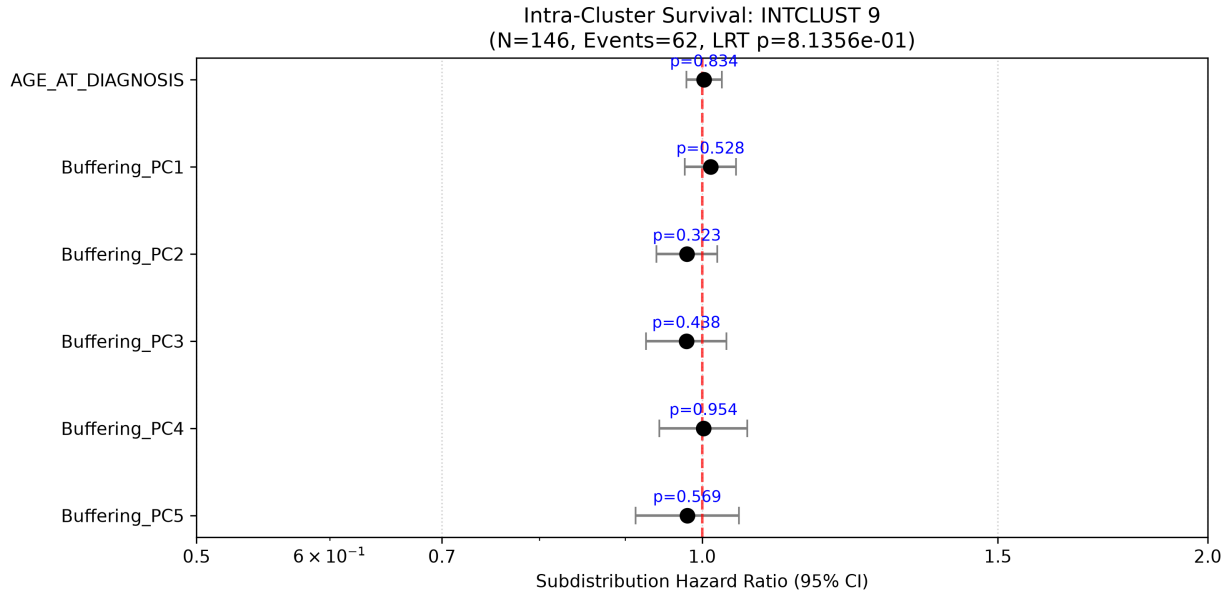


Figure 14: Forest plot displaying the Subdistribution Hazard Ratios (95% CI) for disease-specific mortality within the `INTCLUST_9` subtype ($N = 146$, 62 events) evaluated via a Fine-Gray competing risks model. The plot illustrates the prognostic effects of the top five Buffering State Vectors (`Buffering_PC1` to `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. In contrast to the significant prognostic refinement observed in clusters such as `INTCLUST_4ER+` and `INTCLUST_8`, all features in `INTCLUST_9` cross the null effect line with non-significant p-values. The overall Likelihood Ratio Test (LRT $p \approx 0.814$) confirms that adding these buffering components does not significantly improve upon an age-only baseline model. This demonstrates that while buffering state vectors successfully capture tumor-specific transcriptional rewiring, their prognostic value is heterogeneous and does not universally predict survival across all integrative clusters.

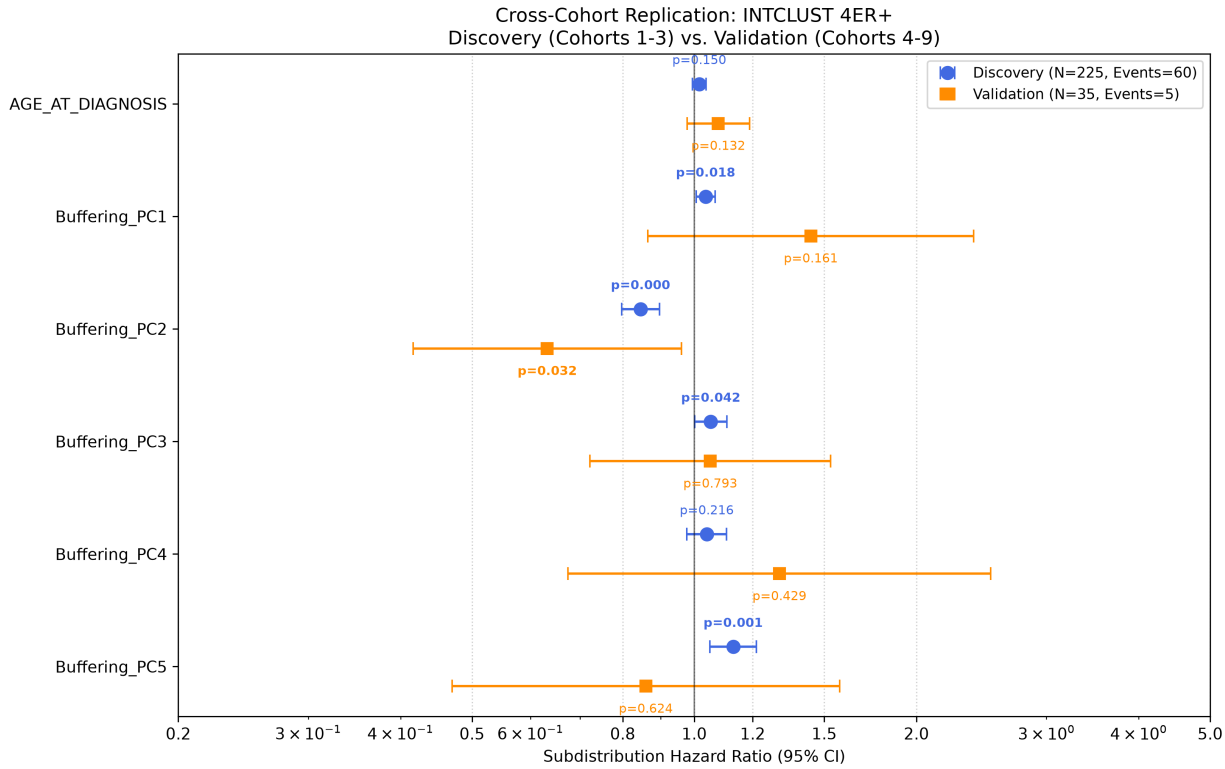


Figure 15: Forest plot detailing the cross-cohort replication of prognostic Buffering State Vectors within the `INTCLUST` 4ER+ subtype. Subdistribution hazard ratios (SHR) and 95% confidence intervals from Fine-Gray competing risks models are shown for the discovery (Cohorts 1-3, blue circles) and validation (Cohorts 4-9, orange squares) sets. The plot illustrates the prognostic value of tumor-specific buffering states for disease-specific mortality, adjusting for `AGE_AT_DIAGNOSIS`. Notably, `Buffering_PC2` demonstrates a robust, statistically significant protective effect that replicates across both cohorts ($p = 0.000$ in discovery; $p = 0.032$ in validation). While other components like `Buffering_PC1` and `Buffering_PC5` show significant associations with increased risk in the discovery set and maintain directional consistency, they do not reach statistical significance in the smaller validation set. These findings confirm that specific transcriptional buffering states, particularly `Buffering_PC2`, represent a reproducible axis of lethality that captures intra-cluster tumor plasticity beyond static genomic classifications.

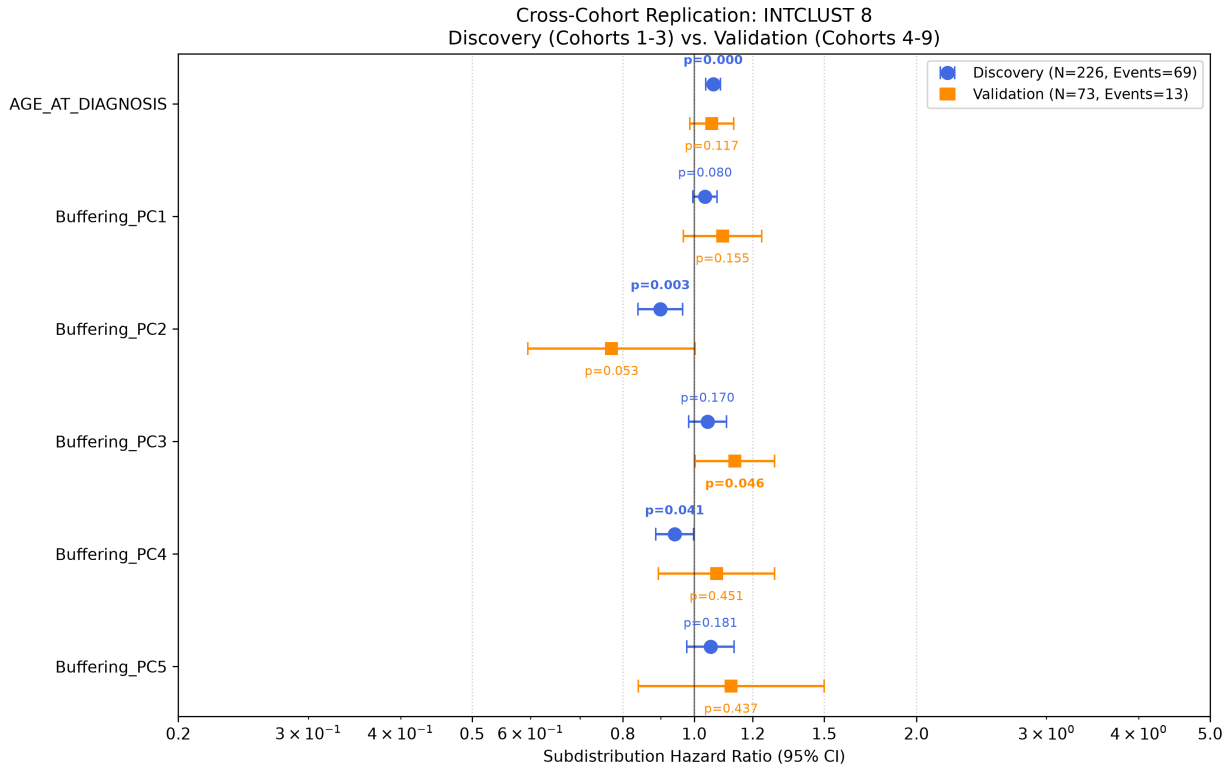


Figure 16: Forest plot illustrating the cross-cohort replication of Buffering State Vectors for disease-specific mortality within the `INTCLUST` 8 subtype. The Subdistribution Hazard Ratios (SHR) and 95% confidence intervals are compared between the discovery (Cohorts 1–3, $N = 226$) and validation (Cohorts 4–9, $N = 73$) sets using Fine-Gray competing risks models adjusted for `AGE_AT_DIAGNOSIS`. The plot demonstrates the prognostic stability of these vectors across the accrual axis; notably, `Buffering_PC2` maintains directional consistency as a protective factor, showing a significant effect in the discovery set ($p = 0.003$) and narrowly missing the significance threshold in the smaller validation set ($p = 0.053$). This indicates that specific transcriptional buffering states provide reproducible intra-cluster prognostic value that is distinct from static copy-number classifications.

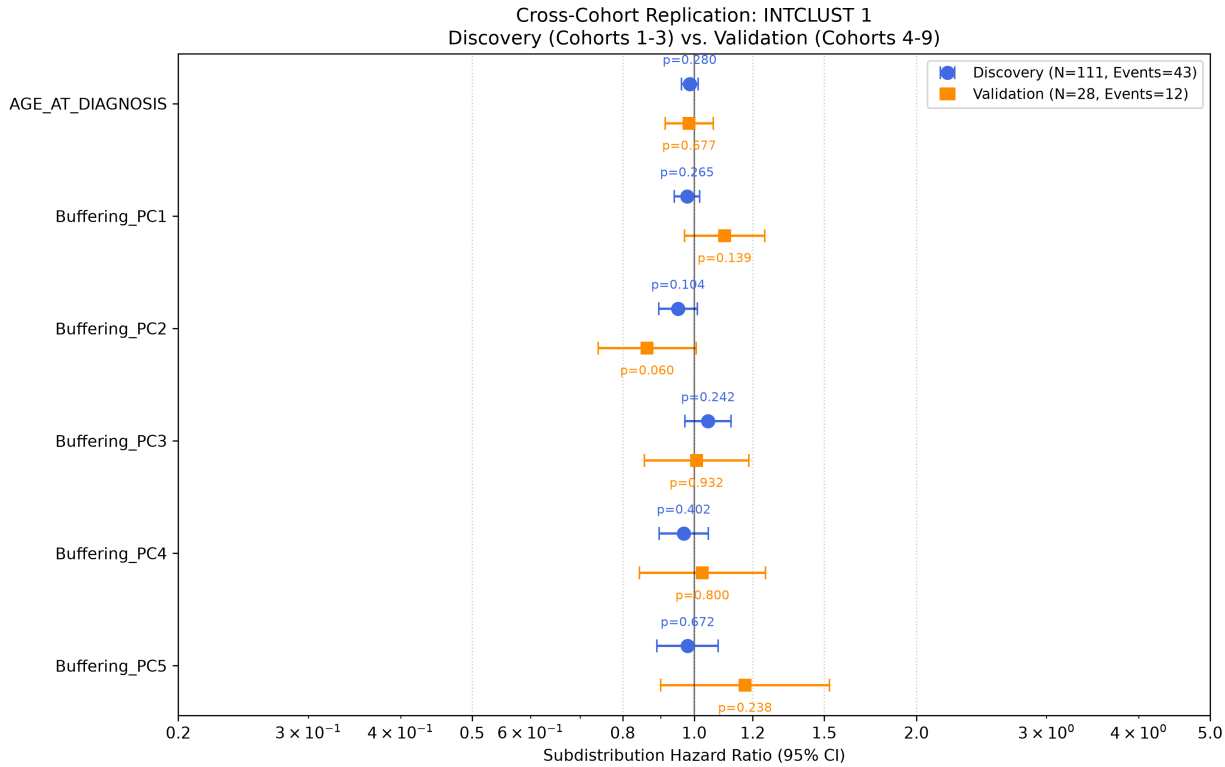


Figure 17: Forest plot detailing the cross-cohort replication analysis of disease-specific mortality for the `INTCLUST` 1 subtype. The plot displays the Subdistribution Hazard Ratios (95% CI) and corresponding p-values derived from Fine-Gray competing risks models for both the discovery (Cohorts 1–3, $N = 111$, 43 events) and validation (Cohorts 4–9, $N = 28$, 12 events) sets. The models evaluate the prognostic utility of the top five Buffering State Vectors (`Buffering_PC1` through `Buffering_PC5`) adjusted for `AGE_AT_DIAGNOSIS`. In contrast to the significant prognostic associations observed in other integrative clusters such as `INTCLUST` 4ER+ and 8, none of the buffering principal components achieve statistical significance ($p > 0.05$) in either the discovery or validation cohorts for `INTCLUST` 1. This indicates that while buffering states quantify tumor-specific transcriptional rewiring, their capacity to significantly refine patient prognosis is heterogeneous and highly subtype-dependent.

18), `INTCLUST` 2 (Figure 19), `INTCLUST` 3 (Figure 20), `INTCLUST` 5 (Figure 21), `INTCLUST` 6 (Figure 22), `INTCLUST` 7 (Figure 23), and `INTCLUST` 9 (Figure 24). Despite these limitations in statistical power for smaller subgroups, the robust replication of `Buffering_PC2` in `INTCLUST` 4ER+ establishes that specific transcriptional buffering states represent a reproducible axis of lethality, capturing critical tumor plasticity beyond static genomic classifications.

3.5 Interpretation of buffering states

These results demonstrate that the buffering of dosage-sensitive hubs—specifically the deviation of mRNA levels from their genomic expectation—constitutes a reproducible axis of lethality within established breast cancer subtypes. The robust replication of `Buffering_PC2` in `INTCLUST` 4ER+ suggests that certain transcriptional states related to DSH buffering are fundamental to the progression of this specific subtype. The fact that these state vectors provide prognostic value within the integrative clusters confirms they are not merely capturing the same signal as existing copy-number-based classifications. Instead, they quantify the plasticity or rewiring of the tumor's response to its own genomic burden. High-risk buffering states (e.g., `Buffering_PC5` in 4ER+) likely reflect a breakdown in dosage-control mechanisms or the activation of alternative transcriptional programs that exacerbate the oncogenic potential of genomic alterations. Conversely,

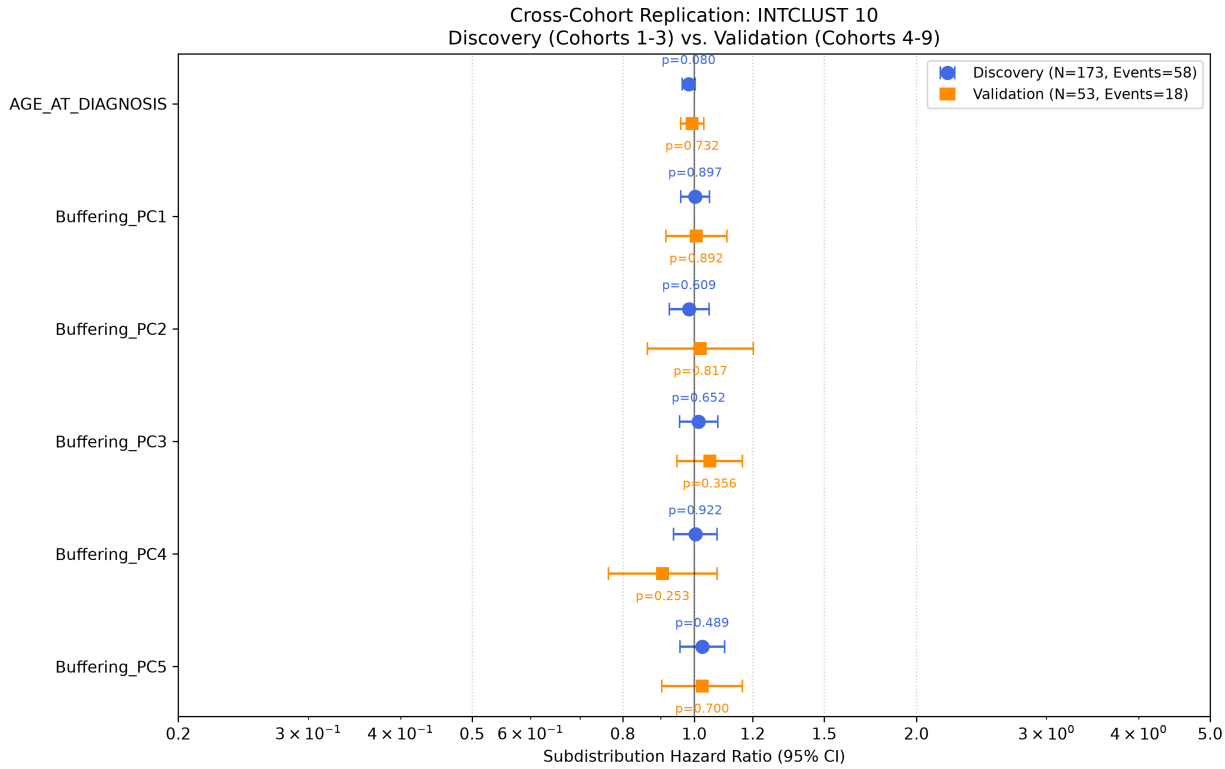


Figure 18: Forest plot displaying the cross-cohort replication analysis of Subdistribution Hazard Ratios (SHR) and 95% confidence intervals for `AGE_AT_DIAGNOSIS` and the top five Buffering State Vectors (`Buffering_PC1` through `Buffering_PC5`) within the `INTCLUST` 10 subtype. Evaluated via Fine-Gray competing risks models for disease-specific mortality, the cohort is divided into a discovery set (Cohorts 1–3, $N = 173$, 58 events) and a validation set (Cohorts 4–9, $N = 53$, 18 events). Unlike the significant prognostic signals found in other integrative clusters, none of the buffering principal components demonstrate a statistically significant association with mortality in `INTCLUST` 10, with all confidence intervals crossing an SHR of 1.0 and p-values exceeding 0.05 in both data splits. This lack of significance indicates that the prognostic value and biological impact of transcriptional buffering states are highly subtype-specific.

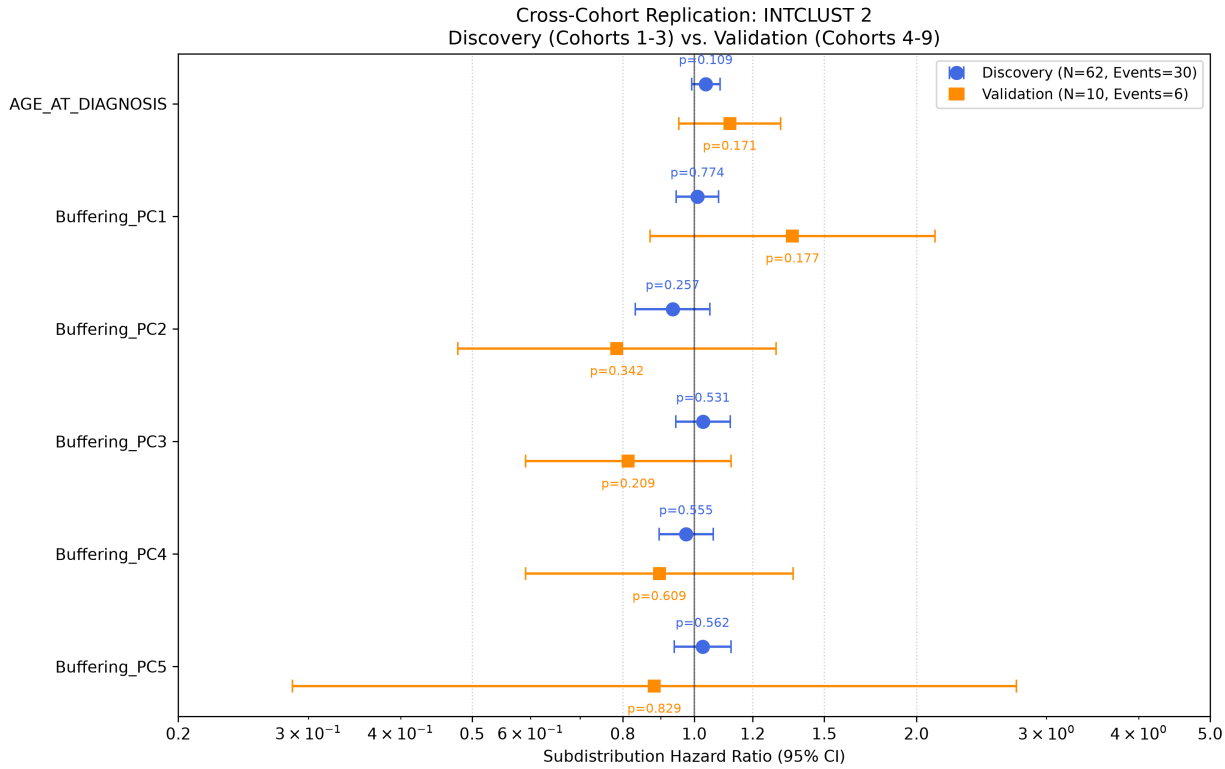


Figure 19: Forest plot illustrating the cross-cohort replication analysis of prognostic effects for Buffering State Vectors and `AGE_AT_DIAGNOSIS` within the `INTCLUST` 2 subtype. The plot compares Subdistribution Hazard Ratios (SHR) and 95% confidence intervals derived from Fine-Gray competing risks models between the discovery set (Cohorts 1-3, $N = 62$, 30 events) and the validation set (Cohorts 4-9, $N = 10$, 6 events). Consistent with the broader findings regarding sample size limitations, the validation set exhibits extremely wide confidence intervals and lacks statistical significance ($p > 0.05$) across all principal components (`Buffering_PC1` through `Buffering_PC5`). This demonstrates that while buffering state vectors can quantify tumor-specific transcriptional rewiring, achieving robust cross-cohort replication is heavily constrained by the small sample sizes and low event counts available in specific validation partitions.

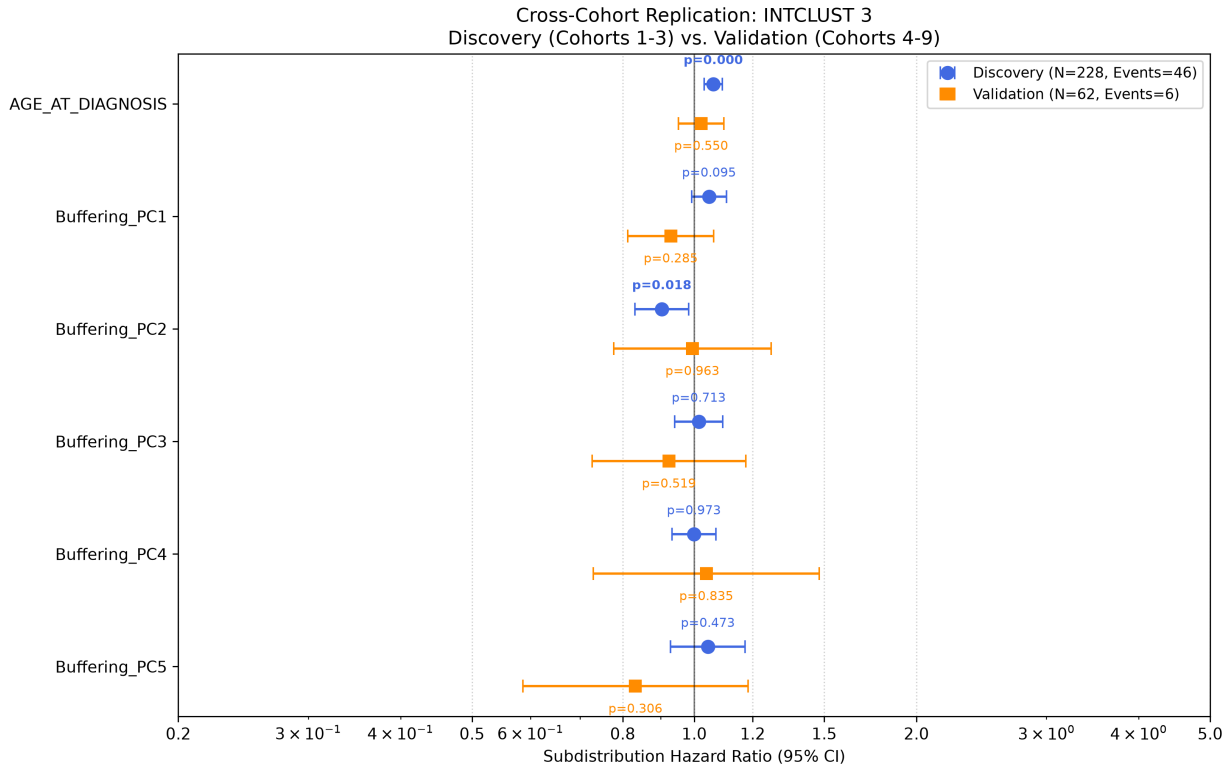


Figure 20: Forest plot illustrating the cross-cohort replication analysis of Buffering State Vectors for disease-specific mortality within the `INTCLUST` 3 subtype. Subdistribution hazard ratios and 95% confidence intervals derived from Fine-Gray competing risks models are compared between the discovery set (Cohorts 1-3, $N = 228$, 46 events) and the validation set (Cohorts 4-9, $N = 62$, 6 events). While `AGE_AT_DIAGNOSIS` and `Buffering_PC2` demonstrate significant prognostic value in the discovery cohort ($p = 0.000$ and $p = 0.018$, respectively), these associations fail to reach statistical significance in the validation set. Consistent with the broader cross-cohort stability analysis, this lack of replication highlights the challenge of validating intra-cluster transcriptional buffering effects when limited by the smaller sample sizes and lower event rates of the validation partition.

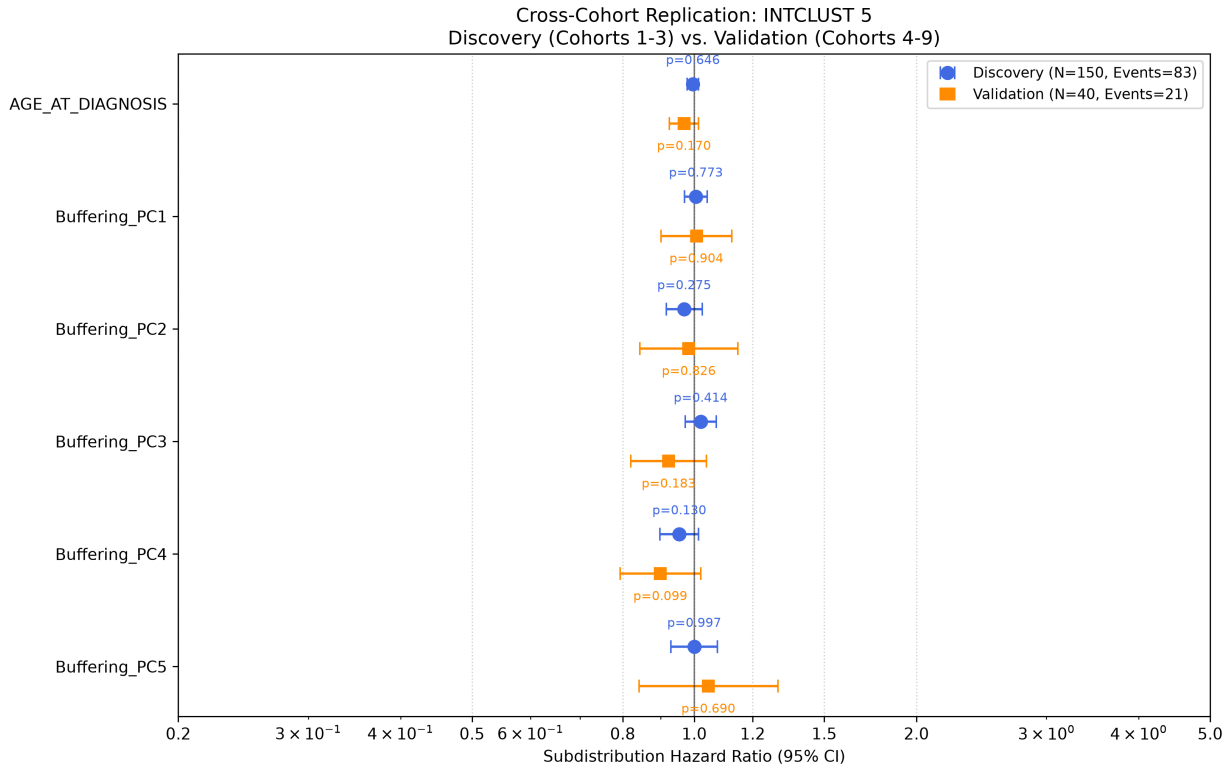


Figure 21: Forest plot illustrating the cross-cohort replication analysis of Buffering State Vectors for predicting disease-specific mortality within the `INTCLUST` 5 subtype. Subdistribution hazard ratios and 95% confidence intervals from Fine-Gray competing risks models are shown for `AGE_AT_DIAGNOSIS` and the first five principal components (`Buffering_PC1` through `Buffering_PC5`). The data is partitioned into a discovery set (Cohorts 1–3, $N = 150$, 83 events) and a validation set (Cohorts 4–9, $N = 40$, 21 events). In contrast to the significant prognostic value observed in other integrative clusters, none of the buffering components achieve statistical significance ($p > 0.05$) in either the discovery or validation sets for `INTCLUST` 5. This demonstrates that the prognostic utility of dosage-sensitive hub buffering is highly subtype-specific and does not provide additional risk stratification within this particular cluster.

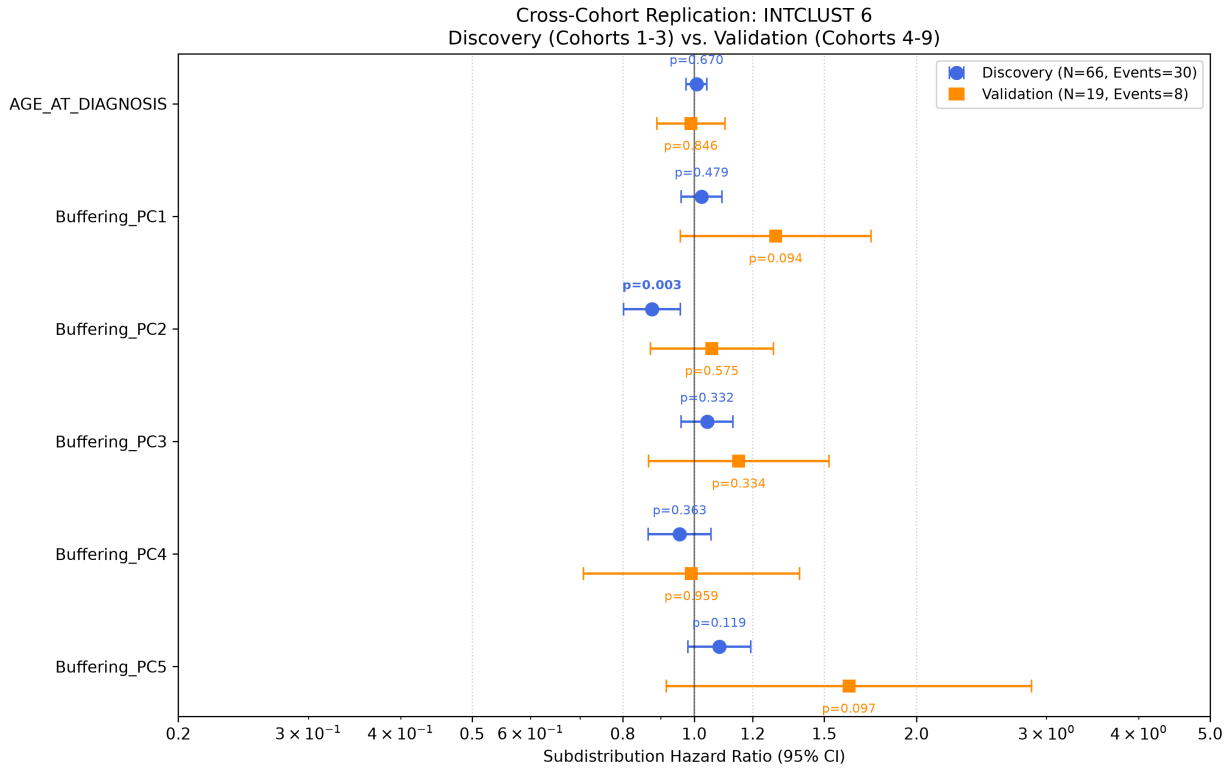


Figure 22: Forest plot illustrating the cross-cohort replication analysis of Buffering State Vectors for predicting disease-specific mortality within the `INTCLUST` 6 subtype. Subdistribution hazard ratios (95% CI) from Fine-Gray competing risks models are compared between the discovery (Cohorts 1-3, $N = 66$) and validation (Cohorts 4-9, $N = 19$) sets for `AGE_AT_DIAGNOSIS` and the top five buffering principal components. Although `Buffering_PC2` shows a statistically significant protective effect in the discovery cohort ($p = 0.003$), it fails to replicate in the validation cohort ($p = 0.575$), exhibiting directional inconsistency. This lack of replication highlights the instability of the prognostic signal within this specific cluster across the accrual axis, which is likely exacerbated by the severely limited sample size and low event count (8 events) in the validation partition.

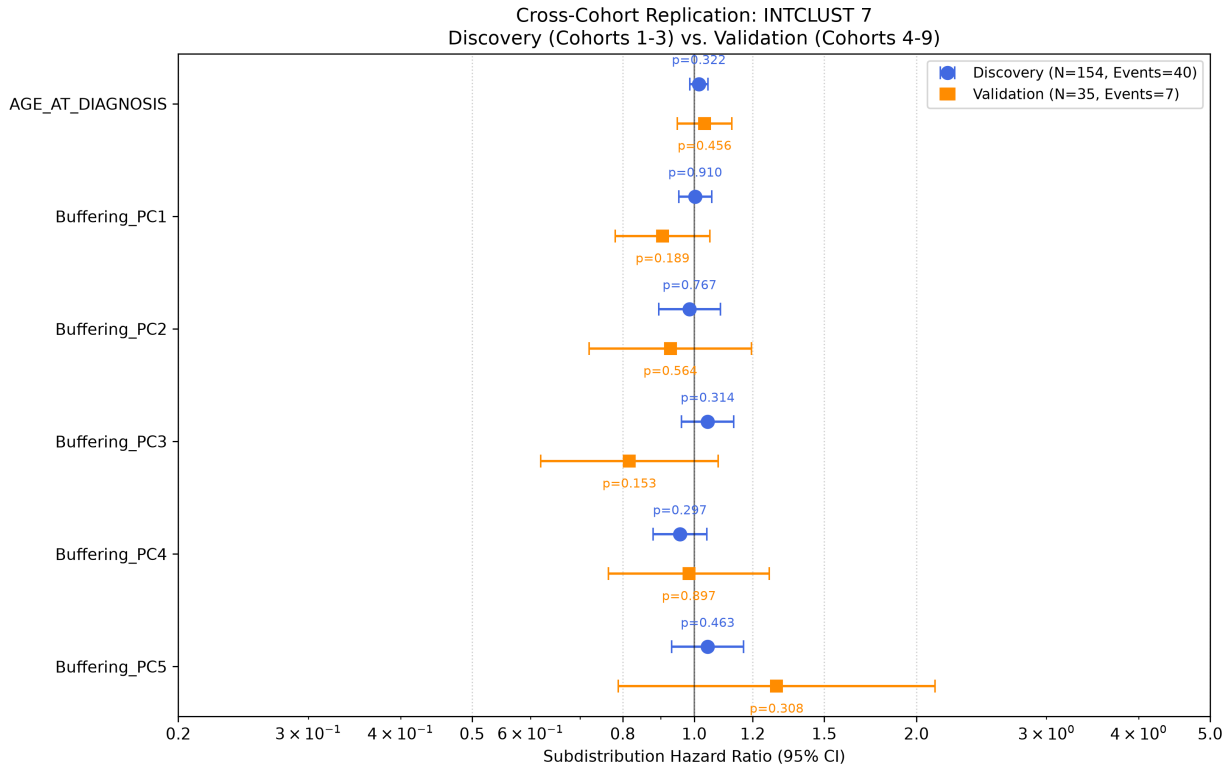


Figure 23: Forest plot displaying the Subdistribution Hazard Ratios (95% CI) for `AGE_AT_DIAGNOSIS` and the top five Buffering State Vectors (`Buffering_PC1` through `Buffering_PC5`) evaluated within the `INTCLUST` 7 subtype. The analysis compares the discovery set (Cohorts 1–3, $N = 154$, 40 events) to the validation set (Cohorts 4–9, $N = 35$, 7 events) using Fine-Gray competing risks models for disease-specific mortality. In contrast to the significant prognostic refinement observed in other clusters, neither age nor the buffering principal components demonstrate statistically significant associations with mortality in this subgroup ($p > 0.05$ for all features). Furthermore, the notably wide confidence intervals in the validation cohort highlight the reduced statistical power associated with its small sample size, illustrating the challenges of cross-cohort replication in smaller partitions.

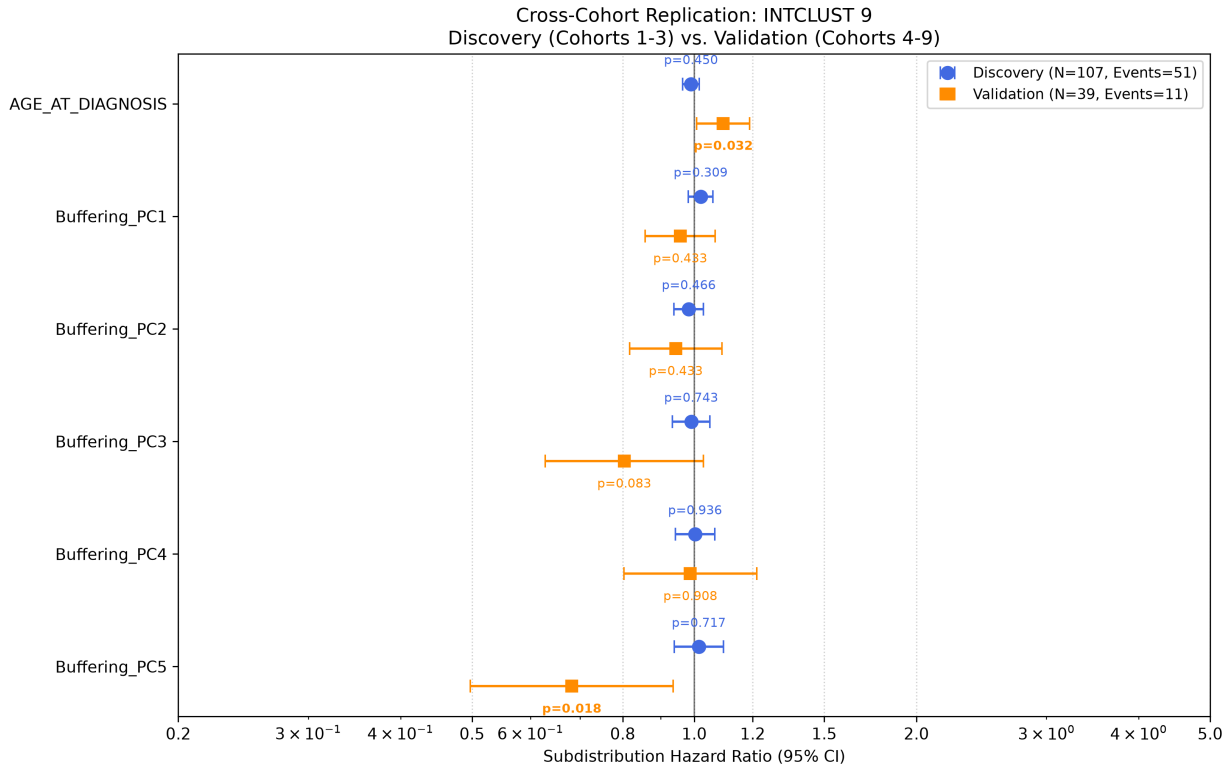


Figure 24: Forest plot illustrating the cross-cohort replication analysis of Buffering State Vectors for the `INTCLUST` 9 subtype. Subdistribution hazard ratios (95% CI) from Fine-Gray competing risks models are compared between the discovery set (Cohorts 1–3, $N = 107$) and the validation set (Cohorts 4–9, $N = 39$). The models evaluate the prognostic impact of age at diagnosis (`AGE_AT_DIAGNOSIS`) and the top five buffering principal components (`Buffering_PC1` through `Buffering_PC5`), which quantify tumor-specific deviations from expected dosage-expression relationships. Unlike the robust signals observed in other clusters, none of the buffering components achieved statistical significance ($p < 0.05$) in the discovery set for `INTCLUST` 9. Although `Buffering_PC5` and age demonstrate significant associations in the smaller validation set, they do not meet the study's strict criteria for cross-cohort replication (which requires prior significance in the discovery set), emphasizing the variability of transcriptional rewiring across different integrative clusters and the challenge of validating prognostic signals in smaller cohort partitions.

protective states (e.g., `Buffering_PC2` in 4ER+) may represent effective buffering where the tumor fails to translate its genomic gains into a transcriptional advantage. This work establishes buffering state vectors as a viable framework for exploring the joint organization of molecular layers and their impact on patient outcomes.

4 Conclusions

In this study, we addressed the challenge of quantifying transcriptional plasticity in breast cancer by introducing a novel analytical framework based on the buffering of dosage-sensitive hubs. While traditional static taxonomies, such as integrative clusters, provide robust genomic stratifications, they inherently obscure the heterogeneous ways individual tumors modulate their genomic burden. By modeling the expected dosage-expression relationships of 500 highly coupled genes and extracting tumor-specific residuals, we constructed multidimensional buffering state vectors. These vectors successfully captured biological variance orthogonal to existing classifications, offering a high-resolution, intra-cluster perspective on transcriptional rewiring.

Our survival analyses demonstrated that these buffering state vectors are not merely descriptive, but carry profound prognostic significance for disease-specific mortality. By employing Fine-Gray competing risks models, we isolated the lethal impact of transcriptional rewiring from competing causes of death. The state vectors significantly refined patient prognosis within specific genomic contexts, most notably the 4ER+ and 8 integrative clusters. We identified distinct transcriptional states, such as the one captured by the second principal component, that confer a strong protective effect against disease-specific death. Conversely, other buffering states were associated with increased mortality risk, highlighting that the lethal consequences of dosage-expression discordance are highly context-dependent and subtype-specific.

A critical strength of this study is the rigorous evaluation of cross-cohort stability. Although validating intra-cluster effects in smaller cohort partitions is inherently limited by statistical power and lower event rates, the protective buffering state identified in the 4ER+ subtype demonstrated robust replication across the accrual axis. This successful validation establishes the transcriptional buffering of dosage-sensitive hubs as a reproducible, independent axis of lethality rather than a statistical artifact of the discovery cohort.

Biologically, these findings suggest that a tumor's capacity to buffer or exacerbate its genomic alterations is a fundamental determinant of its clinical trajectory. Protective buffering states likely represent scenarios where a tumor fails to translate its genomic gains into a transcriptional and oncogenic advantage, effectively neutralizing the potential of its copy number alterations. In contrast, high-risk states may reflect a breakdown in dosage-control mechanisms or the activation of alternative transcriptional programs that allow genomic aberrations to drive aggressive disease phenotypes.

Ultimately, this work demonstrates that evaluating the dynamic, tumor-specific response to genomic copy number alterations provides critical insights missed by static pathway aggregations or global burden normalizations. The buffering state vector framework offers a scalable and biologically interpretable approach to multi-omic integration. By quantifying transcriptional plasticity at the individual tumor level, this methodology holds significant promise for refining breast cancer stratification and uncovering novel, intra-cluster vulnerabilities that drive disease-specific mortality.