Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A therapy-aware multimodal dynamical scaffold reveals residual persistence and relapse-associated escape in pediatric leukemia

Seung-Hwan Kim

seung-hwan.kim@fisher.edu

Fisher College

Article

Keywords:

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10856281/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Pediatric leukemia progression under therapy is not adequately described as either smooth stochastic drift or simple discrete state switching. Instead, response, residual persistence and relapse emerge within therapy-shaped disease states in which malignant programs, regulatory state and ecological context change together. Here we developed a therapy-aware multimodal dynamical scaffold that defines a diagnosis-anchored disease-state reference frame from complementary single-cell transcriptomic, inferred regulatory, ecological and clinical information, and uses an Ornstein–Uhlenbeck–Lévy–branching formulation to interpret longitudinal movement through this space. In the discovery cohort, the scaffold resolved distinct malignant-state basins and microenvironmental contexts, providing an interpretable baseline for longitudinal projection. Across 21 diagnosis-to-relapse transitions, branch-switching trajectories showed larger displacement than branch-continuous trajectories and were enriched in an upper-tail escape subset, supporting a hybrid picture of constrained evolution with punctuated relapse-associated departure. Therapy-aware dynamic summaries further separated three clinically meaningful regimes: response-like constrained states, residual persistent states and escape-prone relapse states. Projection of an independent AML cohort into the frozen scaffold reproduced the treatment-aware organization of end-of-induction/remission and relapse samples, and conservative serial bulk validation preserved the same directional structure at lower resolution. Together, these results define an interpretable precision-oncology map for monitoring therapy response, identifying residual persistence and prioritizing escape-prone disease states in pediatric leukemia.

Introduction

Precision oncology in pediatric leukemia depends not only on identifying baseline molecular subtypes, but also on understanding how disease states reorganize under therapy as patients move from diagnosis to response, residual disease and relapse. Longitudinal cancer evolution can proceed through clonal selection, branching diversification and, in some settings, punctuated change, as shown across hematologic and solid tumors.[1–8] This remains challenging because treatment-linked evolution is shaped simultaneously by malignant-cell state, regulatory circuitry and the surrounding ecological context, whereas many longitudinal models emphasize either gradual stochastic change or discrete state switching in isolation.[3,6,9–16] As a result, clinically important intermediate states, including partial response with residual persistence and early escape-prone reorganization, remain difficult to place within a unified and interpretable framework. With the increasing availability of longitudinal single-cell and multimodal datasets, there is now an opportunity to move beyond static molecular snapshots and build disease-state maps that quantify how pediatric leukemias remain constrained, become destabilized or undergo relapse-associated departure during therapy.[17–22] Such a framework could refine response assessment, identify high-risk residual states and provide a more mechanistic basis for disease monitoring in pediatric leukemia.

Here we address this gap by developing a therapy-aware multimodal dynamical scaffold that defines a diagnosis-anchored disease-state reference frame and quantifies how pediatric leukemia samples move through that space during treatment. The scaffold integrates transcriptomic state, regulatory program activity, ecological or microenvironmental context and clinical phase into a shared latent representation. Longitudinal movement through this reference frame is then interpreted using an Ornstein–Uhlenbeck (OU)–Lévy–branching formulation in which therapy can shift the local attractor, alter effective restoring strength and change the propensity for punctuated departure.[23–33] Using a multimodal single-cell discovery cohort with matched diagnosis, on-therapy, end-of-induction/remission and relapse states, together with an independent AML calibration cohort and conservative serial bulk validation, we asked whether favorable response states remain constrained near a diagnosis-defined basin, whether residual states occupy partially retained but non-collapsed positions, and whether relapse is preferentially associated with branch switching and larger escape-like displacement. Conceptually, this study extends prior hybrid OU–branching and ecotype-aware pediatric leukemia analyses by incorporating therapy-aware modulation and a jump-capable component within a multimodal scaffold.[34–36]

Results

Study design and construction of a therapy-aware multimodal disease-state scaffold

To examine pediatric leukemia evolution in a framework that remains interpretable across treatment phases, we organized the study around a longitudinal design spanning diagnosis (DX), on-therapy response states, end-of-induction/remission (EOI/REM), and relapse (Fig. 1A). Discovery, calibration and validation components served distinct roles. The discovery component provided multimodal single-cell measurements with matched temporal structure, enabling construction of a diagnosis-anchored disease-state scaffold and projection of later treatment phases into the same reference frame. Independent cohorts were then used to assess whether the learned structure generalized beyond the discovery setting, first through external AML single-cell calibration and then through a lower-resolution serial bulk cohort used as a conservative translational validation. This design addressed a central clinical question: whether post-diagnosis leukemia states can be interpreted not as isolated molecular snapshots, but as positions and movements within a shared therapeutic landscape that separates response-like contraction, residual persistence and relapse-linked escape.

We next defined the scaffold by integrating multiple axes of biological information relevant to treatment response and relapse biology (Fig. 1B). Transcriptomic measurements captured expression programs and cell-state composition, regulatory features summarized chromatin- and transcription factor–linked state, ecological context represented immune–stromal and niche organization, and treatment phase encoded the evolving clinical context in which these states were observed. Rather than analyzing these modalities separately, we used them jointly to construct an integrated latent disease-state representation that could serve as a common coordinate system for longitudinal interpretation.

The dynamic model was built around the premise that therapy does not simply add noise to leukemia evolution but reshapes the local landscape through which disease states move (Fig. 1C). In this formulation, treatment phase can modulate attractor location, effective restoring strength, stochastic diffusion and jump propensity, allowing the same scaffold to represent constrained baseline behavior, therapy-associated state retention and punctuated relapse-linked departure within a unified framework. Continuous motion within a local basin was interpreted as mean-reverting stochastic evolution, occasional larger departures as Lévy-like jump behavior, and branch structure as divergence into non-equivalent evolutionary routes.

The analysis workflow proceeded from preprocessing and multimodal integration to latent-state construction and patient-level trajectory reconstruction, followed by therapy-aware dynamic-summary estimation, branch-level analysis, and external projection into the frozen scaffold (Fig. 1D). This architecture was designed to preserve biological interpretability while allowing response, residual and relapse states to be compared directly with the diagnosis-defined baseline.

Finally, the scaffold was organized around three clinically interpretable dynamic regimes inferred from these model components (Fig. 1E). Response-like constrained states remain anchored near a diagnosis-like basin, residual persistent states retain disease structure in partially shifted configurations, and relapse escape states occupy more displaced, branch-switching or jump-sensitive positions. On this basis, we next asked whether the diagnosis-anchored scaffold recovered structured biological organization suitable for longitudinal precision-oncology analysis.

A diagnosis-anchored latent scaffold captures malignant-state structure, ecological context, and regulatory variation

We next asked whether a pretreatment, diagnosis-anchored latent scaffold could recover structured biological organization suitable for longitudinal interpretation. To establish a common reference frame that was not already distorted by therapy-associated movement, we first constructed the scaffold from diagnosis-phase observations in the discovery cohort and examined how clinical timepoint structure was represented within this baseline space (Fig. 2A). This design allowed subsequent EOI/REM and relapse states to be interpreted as displacements from a shared pretreatment configuration rather than as isolated post hoc clusters. In the resulting latent representation, diagnosis samples occupied a coherent but heterogeneous manifold, indicating that the baseline disease-state space retained substantial biological diversity while remaining suitable as a common coordinate system for later longitudinal projection.

We then examined whether this diagnosis scaffold preserved clinically relevant inter-patient and disease-subgroup structure (Fig. 2B). Coloring the latent space by patient or subgroup showed that the geometry was not random with respect to sample identity but instead captured broad disease-state organization distributed across partially overlapping regions of the manifold. This pattern is consistent with a scaffold that retains patient-level and subgroup-level heterogeneity without collapsing all baseline leukemias into

a single undifferentiated state. At the same time, the partial overlap across regions suggests that the latent space captures shared biological structure rather than simply recapitulating sample labels. For a precision-oncology framework, this balance is important: the scaffold must preserve clinically meaningful individuality while still supporting a shared disease-state map onto which future treatment phases can be projected.

We next asked whether ecological context was embedded jointly with malignant-state structure within the same diagnosis-defined reference frame (Fig. 2C). Ecological or microenvironmental contexts occupied non-random and partially distinct territories across the scaffold, indicating that baseline leukemia state was organized in part by surrounding immune and stromal composition rather than by malignant features alone. Thus, ecological variation was not external to the disease-state map but constituted one of its organizing dimensions. This joint representation is important for the later interpretation of therapy response and relapse, because it enables longitudinal movement to be evaluated not only in terms of malignant-state displacement, but also in relation to the ecological landscape in which that movement occurs.

Regulatory program activity added a further, continuous axis of organization across the same latent space (Fig. 2D). Rather than appearing as a purely discrete partition, regulatory signal varied smoothly across the manifold, forming broad gradients that traversed multiple regions of the scaffold. This observation supports the view that the diagnosis landscape contains both basin-like structure and continuous internal variation, consistent with a disease-state framework in which leukemias can remain constrained within a local neighborhood while still exhibiting meaningful regulatory heterogeneity.

When branch assignments were superimposed on the scaffold, the latent space resolved major divergence basins with HSC-like, progenitor-like, GMP-like and Mono/DC-like identities (Fig. 2E). These branch labels provided coarse-grained disease-state identities for subsequent longitudinal analyses, including branch-continuous and branch-switching transitions. Integrating these views, the diagnosis scaffold defined an interpretable pretreatment disease landscape in which malignant-state structure, ecological context and regulatory variation jointly organized baseline leukemia states (Fig. 2F). This scaffold therefore provided the reference frame from which later response-like contraction, residual persistence and relapse-associated escape could be quantified.

Longitudinal projection reveals group-level contraction with an upper-tail subset of relapse-associated departures

We next projected longitudinal samples from the discovery cohort into the frozen diagnosis-anchored scaffold to ask how pediatric leukemia states move through the shared disease landscape during treatment and relapse. Patient-level trajectories showed that longitudinal evolution was neither uniformly progressive nor randomly dispersed across the latent space (Fig. 3A). Instead, individual courses traced heterogeneous but interpretable paths relative to the diagnosis-defined baseline, with some intervals remaining comparatively constrained and others departing more strongly from their initial positions. This

projection framework made it possible to compare post-diagnosis states within a common coordinate system and to distinguish modest longitudinal movement from larger therapy-associated reorganization.

At the group level, the projected trajectories supported a hybrid pattern of partial contraction together with relapse-associated displacement (Fig. 3B). EOI/REM states remained linked to the diagnosis scaffold rather than forming a completely separated configuration, consistent with retention of disease-state structure during favorable or partially favorable treatment response. By contrast, relapse samples occupied a broader and more displaced region of the latent space, and the group median shifted away from the diagnosis median. Thus, the dominant longitudinal picture was not one of uniform collapse into a single response attractor, but one in which some trajectories remained constrained while others moved into more divergent relapse-associated positions.

We then quantified this divergence by measuring high-dimensional DX→REL displacement for each longitudinal transition. Branch-switching intervals were shifted toward larger DX→REL displacement, whereas branch-continuous intervals were concentrated at lower values (Fig. 3C). Importantly, the most extreme departures were not distributed evenly across all trajectories. Instead, the upper tail above the empirical DX→REL threshold was enriched for branch-switching transitions, indicating that the largest longitudinal movements were preferentially associated with changes in branch identity rather than movement within the same branch basin. These results support the interpretation that relapse-associated escape is concentrated in a restricted subset of trajectories rather than representing the default behavior of all recurrent disease.

To further rank these intervals, we decomposed a relapse-escape score into a displacement component and an added branch-switching contribution (Fig. 3D). This analysis again highlighted a compact highest-scoring subset composed predominantly of branch-switching intervals, including transitions from GMP-like, HSC-like and Mono/DC-like starting states into alternative relapse basins. In contrast, branch-continuous intervals generally occupied lower portions of the score distribution even when they exhibited measurable displacement.

Taken together, longitudinal projection into the diagnosis scaffold resolved two coexisting features of pediatric leukemia evolution: a broader background of constrained or moderately displaced trajectories, and an upper-tail subset of branch-switching, relapse-associated departures most consistent with escape-like disease reorganization (Fig. 3E). More broadly, the coexistence of constrained trajectories and an upper-tail escape subset is consistent with models of cancer evolution in which gradual change and punctuated departure coexist rather than exclude one another.[6,7]

Therapy-aware dynamical parameters separate response-like, residual persistent, and escape-prone disease regimes

We next asked whether interpretable therapy-aware dynamic summaries derived from the scaffold could distinguish clinically meaningful disease regimes beyond geometric displacement alone. Given the

sparse longitudinal sampling, these quantities were treated as descriptive summaries of latent disease-state dynamics rather than as direct estimates of continuous-time stochastic parameters. We therefore quantified how strongly projected states remained restored toward a local attractor, how far the inferred attractor shifted from the diagnosis baseline, and how much instability or jump-like escape propensity was present within a given interval (Fig. 4A–D). This analysis was motivated by a precision-oncology question: whether favorable treatment response, residual persistence, and relapse-associated escape occupy measurably different dynamic regimes that could support monitoring and triage. Rather than assuming that all non-diagnosis states reflect the same biology, the therapy-aware framework tests whether post-diagnosis samples differ in the balance between residual constraint and escape-like reorganization.

The first informative axis was effective restoring strength. Diagnosis-phase states were generally compatible with a constrained baseline regime, whereas relapse-phase states showed a broader distribution of effective restoring behavior, with some relapse samples retaining partial constraint and others occupying more weakly stabilized positions (Fig. 4A). The limited number of EOI/REM observations in this analysis was treated as exploratory, but these samples illustrate that therapy-linked residual states need not collapse into a maximally restored basin. Instead, they can remain biologically organized in partially shifted positions relative to the diagnosis scaffold. This is an important distinction for clinical interpretation, because a residual state that remains structured but incompletely normalized is different from a fully contracted response state and may still retain the substrate for subsequent progression.

A second informative axis was attractor displacement from the diagnosis baseline. Relapse samples tended to occupy more displaced attractor positions than diagnosis samples, while exploratory EOI/REM states could remain at intermediate or retained locations rather than returning completely to the baseline attractor (Fig. 4B). In other words, post-therapy disease organization was not adequately described as simple restoration toward the pretreatment state. Some samples remained relatively close to the diagnosis-defined basin, consistent with response-like or constrained behavior, whereas others persisted in shifted positions that are more consistent with retained residual disease structure. This diagnosis-referenced interpretation is central to the scaffold: disease states are evaluated not in isolation, but in terms of how far their inferred dynamical center has moved from the pretreatment landscape.

We then examined effective instability and jump propensity as complementary indicators of escape-like behavior. Overall instability showed substantial overlap between diagnosis and relapse, arguing against a simplistic model in which relapse is uniformly more diffusive than baseline disease (Fig. 4C). By contrast, jump-sensitive scoring was more discriminative at the level of interval class: branch-switching intervals showed significantly higher jump scores than branch-continuous intervals (Fig. 4D), indicating that punctuated transition behavior is concentrated in the subset of trajectories that also change branch identity. This result strengthens the interpretation from the longitudinal projection analysis by showing that the upper-tail relapse-associated departures are not only larger in magnitude, but also dynamically distinct in a way consistent with escape-prone reorganization.

Together, these parameters resolved three clinically meaningful dynamic regimes summarized schematically in Fig. 4E. A baseline response-like regime remained anchored near the diagnosis attractor and was characterized by stronger effective restoration and lower punctuated-transition propensity. A residual persistence regime captured states that were not maximally displaced from the diagnosis baseline yet remained incompletely contracted and could preserve weakly constrained malignant structure after therapy. Finally, an escape-prone relapse regime was enriched for branch switching, larger diagnosis-to-relapse displacement, and elevated jump-like behavior. These regimes are not intended as rigid clinical labels, but as interpretable dynamical categories that organize longitudinal leukemia states in a way that is directly relevant to precision oncology: distinguishing favorable contraction from retained residual structure and from the smaller subset of disease states most consistent with relapse-associated escape.

Branch structure and ecological context define non-equivalent routes to relapse-associated escape

We next asked whether relapse-associated escape arose uniformly across the diagnosis scaffold or instead depended on the branch basin and ecological context from which a trajectory originated. When DX→REL intervals were stratified by branch identity, the resulting transition map showed both branch-continuous and branch-switching routes across the 21 matched intervals (Fig. 5A). Relapse could arise from multiple diagnosis basins, including HSC-like, progenitor-like, GMP-like and Mono/DC-like states, but the transition patterns were not interchangeable. Some starting branches more often preserved branch continuity, whereas others contributed disproportionately to branch-switching trajectories. This indicated that relapse-associated movement through the scaffold was branch-structured rather than a uniform shift toward a single relapse state.

We then examined whether these branch basins differed in ecological or microenvironmental context. Ecotype composition varied across the major branch states, with each branch showing a distinct mixture of ecological backgrounds (Fig. 5B). Thus, branch identity was not simply a malignant-state label; it was embedded within a broader microenvironmental context. This observation supports the interpretation that relapse-associated movement through the scaffold reflects coordinated malignant and ecological organization rather than cell-intrinsic state change alone.

To characterize these branch basins more directly, we summarized scaffold program composition within each branch. The branch-level heat map showed that HSC-like, progenitor-like, GMP-like and Mono/DC-like basins retained distinct program signatures across malignant-state and auxiliary scaffold programs (Fig. 5C). These differences reinforce the view that the diagnosis scaffold is not partitioning samples arbitrarily but instead recovers biologically interpretable disease-state regions. A branch switch at relapse therefore represents movement into a region with a different composite malignant, ecological and programmatic identity.

Finally, we asked whether escape propensity differed by starting branch. Branch-level jump-sensitive scores showed that the diagnosis basin from which a trajectory began was associated with different average escape tendencies (Fig. 5D). GMP-like starting states showed the highest mean jump-sensitive score and a high branch-switching fraction among the available intervals, consistent with a particularly escape-prone configuration, although this subset was small and should be interpreted cautiously. HSC-like and Mono/DC-like starting states showed more intermediate escape behavior, whereas progenitor-like starting states were comparatively lower on the jump-sensitive axis. These contrasts suggest that relapse-associated escape is branch-conditioned: some diagnosis basins appear more permissive to punctuated departure, whereas others more often support branch continuity or less extreme movement.

Together, these analyses show that relapse-associated escape follows non-equivalent routes defined jointly by branch structure, ecological context and scaffold program composition. The diagnosis scaffold therefore does more than separate baseline disease states; it identifies starting regions that differ in their microenvironmental embedding and escape propensity. This branch- and ecology-conditioned view provides a mechanistic bridge between the longitudinal displacement results and the clinical interpretation of escape-prone disease states.

External AML projection reproduces treatment-aware organization in an independent cohort

We next asked whether the treatment-aware organization learned in the discovery cohort generalized to an independent pediatric AML dataset rather than reflecting cohort-specific structure alone. To test this, we projected the external AML cohort into the frozen diagnosis-anchored scaffold without relearning the latent space (Fig. 6A). This provided a stringent calibration step for the framework, because successful projection required the scaffold to capture disease-state organization that remained meaningful outside the original discovery cohort. In the projected space, external samples aligned with the discovery-defined reference field rather than dispersing randomly, indicating that the diagnosis-anchored scaffold preserved transferable structure across cohorts.

We then examined whether clinical timepoints in the external cohort retained the treatment-aware ordering observed in the discovery data. When projected samples were colored by phase, DX, EOI/REM and REL states occupied related but non-identical regions of the scaffold (Fig. 6B). EOI/REM samples remained linked to the broader diagnosis-defined landscape, whereas relapse samples tended to occupy more displaced positions. Importantly, the external cohort contained fully longitudinal trajectories, allowing direct visualization of time-resolved movement through the frozen scaffold. These trajectories traversed the same reference frame used for the discovery cohort rather than requiring a cohort-specific embedding, supporting the view that the scaffold captures reusable therapeutic organization rather than descriptive structure limited to one dataset.

To quantify this calibration, we compared attractor displacement from the discovery DX baseline across external timepoints. External EOI/REM samples remained close to the diagnosis attractor on average,

with displacement values comparable to those of external DX samples, whereas relapse samples were more displaced than EOI/REM samples (Fig. 6C). At the same time, EOI/REM states did not collapse into a single tightly restored cluster. Instead, they remained diffusely organized within the projected space, consistent with the discovery-cohort interpretation that residual or response-like states can preserve structured biological variation even when they remain relatively close to the diagnosis-defined basin. This distinction is translationally important: favorable-treatment or residual states need not represent complete reversion to a baseline attractor but can occupy retained positions that are still biologically organized and clinically interpretable.

The external cohort therefore reproduced the central treatment-aware logic of the scaffold while reinforcing its nuanced interpretation (Fig. 6D). The independent data supported three linked conclusions: EOI/REM states can remain near the diagnosis-defined attractor, relapse states are more displaced than EOI/REM states, and residual-like states can remain diffusely organized rather than collapsing into a uniformly contracted endpoint. Because these conclusions were obtained by projection into a frozen discovery scaffold, they support the transferability of the disease-state geometry rather than overfitting to one cohort. We therefore interpret the external AML analysis as an independent calibration of the scaffold's precision-oncology utility: preserving a diagnosis-referenced map in which treatment phases can be compared directly, residual states can be recognized as structured intermediate configurations, and relapse can be identified as a more displaced disease-state condition.

Clinical translation and serial bulk validation preserve disease-state ordering

We finally asked whether the therapy-aware structure learned from multimodal single-cell data could be rendered as a clinically interpretable summary and remain informative under lower-resolution validation. To address this translational question, we summarized disease states using scaffold-derived dynamic coordinates, focusing on attractor displacement from the diagnosis baseline and effective restoring strength as compact descriptors of response-like constraint, residual persistence and escape-prone departure (Fig. 7A). In this reduced representation, discovery and external AML samples occupied an ordered dynamic map rather than collapsing into an undifferentiated cloud. States with low displacement and stronger restoration localized to a constrained response-like region, states with intermediate displacement and weaker contraction occupied a residual persistent region, and more displaced states localized to an escape-prone relapse-like region.

Representative cases were then rendered as a dynamic-state scorecard linking each sample to scaffold-derived metrics, branch identity, state tier and jump or branch-switching behavior (Fig. 7B). This scorecard illustrates how the scaffold can be translated from a high-dimensional modeling framework into a more interpretable state-summary view. Rather than serving as a fixed clinical classifier, the map and scorecard provide a compact representation of the quantities most relevant to disease monitoring: how far a state has moved from the diagnosis baseline, how strongly it remains constrained, whether it

occupies a branch-continuous or branch-switching trajectory, and whether it shows jump-sensitive escape behavior.

We next tested whether this directional organization was preserved in an external serial bulk RNA-seq cohort, which provides a deliberately lower-resolution and therefore more demanding validation setting. Because bulk measurements do not resolve the same single-cell multimodal structure, we treated this analysis conservatively and asked only whether the major disease-state ordering inferred from the scaffold could still be recovered. Even under this restricted setting, serial bulk samples retained a clinically meaningful ordering across disease classes (Fig. 7C). Control samples were positioned furthest from the leukemia-associated discovery anchor, diagnosis leukemia samples were shifted toward the disease-state region, and response samples showed directional movement away from the diagnosis leukemia distribution. The first response state (r1) showed lower discovery-anchored displacement than diagnosis leukemia samples, consistent with treatment-associated movement away from the baseline leukemia configuration, whereas the later response state (r2) maintained this response-linked directional pattern despite limited sample size.

This lower-resolution validation addresses a central practical issue in precision oncology: whether a framework derived from rich multimodal discovery data can still yield interpretable structure when applied to assays that are more deployable but biologically less resolved. We do not interpret the bulk cohort as a direct replication of all single-cell parameters or branch-specific features. Rather, we interpret it as a conservative test of directional consistency. Under that test, the major logic of the scaffold was preserved: diagnosis leukemia occupied a disease-associated position distinct from control, treatment response moved samples away from that diagnosis-linked state, and response states remained ordered relative to one another rather than becoming randomly distributed.

Taken together with the discovery and external AML single-cell analyses, the serial bulk validation supports a clinically oriented summary of the framework (Fig. 7D). The multimodal scaffold identifies constrained states compatible with favorable response, residual states that remain biologically organized without complete normalization, and escape-prone states associated with larger displacement and relapse-linked reorganization. The bulk validation does not resolve every mechanistic detail of these regimes, but it shows that their directional organization is not confined to the highest-resolution setting. These results support the therapy-aware scaffold as a clinically interpretable disease-state map for monitoring treatment response, recognizing persistent residual structure and prioritizing states most compatible with relapse-associated escape.

Discussion

In this study, we developed a therapy-aware multimodal dynamical scaffold to interpret pediatric leukemia progression as movement through a clinically meaningful disease-state landscape rather than as a series of disconnected molecular snapshots. The central finding is that longitudinal leukemic evolution follows a hybrid dynamic logic: many trajectories remain partially constrained around

diagnosis-anchored or residual-like states, whereas a smaller but important subset undergoes larger relapse-associated departures enriched for branch switching and elevated jump-sensitive behavior. By integrating malignant-state composition, regulatory activity, ecological context and treatment phase into a shared diagnosis-referenced framework, the scaffold yields interpretable dynamic regimes that align naturally with precision-oncology questions of response, persistence and relapse risk. These findings extend clonal-evolution, ecosystem and state-attractor views of cancer by embedding relapse within a therapy-aware disease-state landscape.[1–16,33] The resulting map is not intended as a purely abstract latent representation; rather, it provides a translational scaffold for distinguishing response-like constraint from residual persistence and escape-prone relapse-associated reorganization, with external cohort projection and lower-resolution validation supporting the broader relevance of this structure.

These results support a view of pediatric leukemia evolution in which treatment response is not adequately described by uniform collapse toward a single post-therapy endpoint. Instead, response-like constrained states, residual persistent states and escape-prone relapse states occupy related but non-equivalent regions of a shared disease-state space. This distinction is important for precision oncology because residual disease that remains biologically organized, even when it is not maximally displaced from diagnosis, may still retain the substrate for later progression. In this setting, geometry alone is insufficient. The combination of attractor displacement, restoring strength, branch continuity and jump-sensitive behavior provides a more informative description of how disease states respond to therapy or move toward relapse. The upper-tail subset of departures is compatible with punctuated evolutionary behavior described in other tumor systems,[6,7] whereas the ecology-conditioned routes to recurrence are consistent with microenvironment-centered models of carcinogenesis and therapeutic response.[9–13]

The branch-structured analyses further suggest that relapse-associated escape is not uniformly distributed across the baseline disease landscape. Distinct diagnosis basins differed in ecological backdrop, scaffold-program composition and escape propensity, indicating that the route taken through the scaffold may carry biological and potentially clinical significance. We interpret these branch- and ecology-conditioned routes conservatively. They do not imply deterministic lineage conversion or complete mechanistic resolution of relapse biology. Rather, they show that the diagnosis-anchored scaffold preserves enough structure to distinguish trajectories that remain relatively constrained from those that undergo more substantial and punctuated reorganization. This ability to separate non-equivalent routes to recurrence is one of the main translational strengths of the framework.

Several limitations should be acknowledged. First, some longitudinal subgroups, particularly EOI/REM and certain branch-conditioned subsets, were modest in size, so their interpretation should emphasize effect direction, relative ordering and reproducible geometric structure rather than over-precise parameter estimation. Second, the serial bulk RNA-seq validation was intentionally conservative and therefore cannot reproduce the full multimodal single-cell geometry or branch-level detail captured in the discovery analysis. Third, the OU–Lévy–branching formulation is used here as an interpretable summary of latent disease-state dynamics rather than as definitive identification of a continuous-time jump process or a complete mechanistic account of the underlying cellular biology.[23–32] Even with these

limitations, the framework showed coherent structure across discovery, external AML calibration and lower-resolution validation, supporting its value as a precision-oncology map for therapy response, residual persistence and relapse-associated escape.

Prospective validation in longitudinal clinical cohorts will determine whether scaffold-derived dynamic states improve prediction of measurable residual disease, relapse timing, or therapy response beyond conventional molecular classification. More densely sampled longitudinal single-cell, spatial and multimodal datasets will also be needed to estimate therapy-dependent attractor shifts, branch-transition probabilities and jump propensity with greater temporal resolution. In this sense, the present study provides a clinically interpretable framework and empirical scaffold rather than a final mechanistic model: it defines how therapy-aware ecological dynamics can be mapped, validated and translated into response monitoring and relapse-escape prioritization in pediatric leukemia.

Methods

Study design and analytical overview

We designed the study to determine whether longitudinal pediatric leukemia samples could be organized within a shared, diagnosis-anchored disease-state framework that distinguishes response-like constraint, residual persistence and relapse-associated escape under therapy. The analysis used a multimodal single-cell discovery cohort to construct the baseline latent scaffold and define the primary therapy-aware dynamical structure, an independent pediatric AML cohort for external calibration by frozen-space projection, and a serial bulk RNA-seq leukemia cohort for conservative cross-modality validation. Across cohorts, the central analytical objective was to interpret post-diagnosis samples not as isolated timepoint-specific clusters, but as positions and movements within a common therapeutic landscape referenced to pretreatment disease organization.

The analytical framework integrated transcriptomic state, regulatory-state information, ecological or microenvironmental context and treatment phase into a shared latent representation. Longitudinal movement through this space was summarized using a therapy-aware Ornstein–Uhlenbeck (OU)–Lévy–branching formulation. This design allowed us to evaluate whether favorable treatment states remained constrained near a diagnosis-defined basin, whether residual states occupied partially retained positions, and whether relapse was enriched for larger displacement, branch-switching behavior and jump-sensitive escape signatures.

Public datasets and cohort roles

Four public pediatric leukemia datasets were assigned predefined analytical roles. The primary discovery cohort was GSE235063, a longitudinal pediatric AML single-cell atlas of treatment response used here for scaffold construction, branch analysis and primary therapy-aware dynamical inference.[20] Raw-data processing and per-sample object construction for the discovery workflow are summarized in Supplementary Fig. S1 and indexed in Supplementary Data 1. Diagnosis-state scaffold variables,

baseline ecological summaries and branch assignments for this cohort are provided in Supplementary Data 2, with baseline state-space heterogeneity and covariance summarized in Supplementary Fig. S3.

The independent AML calibration cohort was GSE235923, a pediatric AML longitudinal cohort used for external projection into the frozen discovery scaffold.[21] Cross-cohort calibration of projected ecotype and malignant-state summaries is shown in Supplementary Fig. S5, and external AML projection outputs and dynamic-parameter summaries are provided in Supplementary Data 5. GSE227122 was used for cross-lineage transfer and contextual validation of the frozen ecological scaffold, including evaluation of whether projected timepoint organization remained interpretable outside the primary AML discovery setting.[22] Strict cross-lineage transfer results and support metrics are summarized in Supplementary Fig. S7 and included in Supplementary Data 5. Conservative cross-modality validation was performed using GSE163634, a serial bulk RNA-seq leukemia cohort used to test whether the major directional organization inferred from the scaffold remained detectable under a lower-resolution assay.[8] Leukemia-versus-control distributions of transferable validation variables are shown in Supplementary Fig. S2, and serial bulk validation outputs are provided in Supplementary Data 6.

Clinical annotations were harmonized into broad phase labels used throughout the analysis: diagnosis (DX), on-therapy or treatment-linked response states, end-of-induction or remission (EOI/REM), and relapse (REL). Because original study designs differed across cohorts, phase harmonization was performed at the level of clinically interpretable treatment stage rather than exact protocol-day matching. Matched DX→REL displacement summaries, branch-continuous and branch-switching annotations, relapse-escape scores, jump-candidate rankings and model-comparison outputs are provided in Supplementary Data 3 and Supplementary Data 6, with supporting displacement and heavy-tail diagnostics shown in Supplementary Fig. S6. Therapy-aware dynamic parameters and state-tier assignments are provided in Supplementary Data 4. No new patient recruitment was performed for this study; all analyses were conducted on publicly available de-identified data.

Discovery-cohort preprocessing and sample-level state definition

For the discovery cohort, processed single-cell matrices and accompanying metadata were imported into a standardized computational workflow. Raw-data processing, archive extraction, per-sample object construction and cohort-level summary generation are described in Supplementary Note 1 and summarized in Supplementary Fig. S1. Cells failing basic quality-control criteria, lacking adequate annotation support or not retained in the finalized preprocessing pipeline were excluded before downstream analysis. The primary analytical unit was the sample rather than the individual cell, because the main goal was to reconstruct longitudinal disease-state movement at the patient or interval level. Sample-level metadata, preprocessing summaries and inclusion information are provided in Supplementary Data 1.

For each retained sample, we derived complementary feature blocks capturing malignant-state structure, regulatory-state variation and ecological context. Transcriptomic state was summarized using scaffold-relevant expression features and composition-aware malignant-state descriptors. Regulatory state was represented using program-level or chromatin-linked activity summaries aligned to the same sample-level structure, motivated by prior work on single-cell multi-omics, quantitative gene-regulatory modeling and mixed-phenotype leukemia regulatory programs.[17–19] Ecological context was quantified using sample-level microenvironmental composition, including non-malignant and malignant-associated contextual proportions, informed by recent longitudinal pediatric AML single-cell studies.[20–22] All feature blocks were standardized before multimodal integration so that no single modality dominated the latent representation because of scale alone. Diagnosis-state variables, ecotype summaries, branch assignments and baseline scaffold coordinates are provided in Supplementary Data 2, with leukemia-versus-control validation-variable distributions and baseline covariance diagnostics shown in Supplementary Figs. S2 and S3.

Here, “multimodal” refers to the integration of complementary transcriptomic, regulatory-program, ecological-composition and clinical-phase feature blocks within a shared disease-state representation; it does not imply that every feature block was measured using a distinct experimental assay in every cohort.

Construction of the diagnosis-anchored multimodal latent scaffold

To define a baseline disease-state reference frame, the latent scaffold was constructed using diagnosis-phase samples from the discovery cohort only. Restricting scaffold learning to pretreatment observations ensured that the latent manifold reflected baseline disease organization rather than a mixture already distorted by treatment-associated movement. Later EOI/REM and relapse samples were therefore analyzed by projection into a fixed diagnosis-defined space rather than by co-embedding all timepoints simultaneously. Baseline diagnosis-state variables, scaffold coordinates and covariance summaries are provided in Supplementary Data 2 and summarized in Supplementary Fig. S3.

The integrated scaffold combined transcriptomic, regulatory and ecological feature blocks into a reduced latent representation at the sample level. We retained a low-dimensional representation suitable for visualization, branch assignment and longitudinal projection while preserving broad patient-level structure, ecological organization and continuous regulatory gradients. Reduced equilibrium diagnostics and leave-one-out summaries supporting the interpretability of the scaffold axes are shown in Supplementary Fig. S4. This diagnosis-anchored latent space served as the common coordinate system for all subsequent trajectory, branch and therapy-aware dynamical analyses, including DX→REL displacement analysis, branch-continuous and branch-switching classification, external AML projection and cross-lineage transfer (Supplementary Data 3–5).

Ecological-context inference

Ecological context was derived from sample-level summaries of the surrounding microenvironment. Cell-type fractions or analogous compositional descriptors were computed within each sample and standardized across the discovery cohort. These ecological features were used both as direct inputs to multimodal scaffold construction and as interpretable descriptors of branch-specific context after latent-state learning, consistent with prior bone marrow stromal and AML ecosystem mapping studies.[12,13] Ecotype summaries, ecological principal-component coordinates and branch-associated ecological variables are provided in Supplementary Data 2.

For visualization and downstream interpretation, samples were grouped into recurring ecological contexts, or ecotypes, representing distinct immune or stromal-biological backgrounds. These ecotypes were used descriptively to assess whether baseline branch basins and longitudinal transitions occurred in non-random ecological settings. Branch-specific ecological composition was then summarized to evaluate whether HSC-like, progenitor-like, GMP-like and Mono/DC-like basins differed in microenvironmental context. Cross-cohort calibration of ecotype-derived coordinates is shown in Supplementary Fig. S5, and branch-specific ecological summaries used for the main branch-context analysis are provided in Supplementary Data 2. The ecological layer was therefore treated as part of the disease-state definition itself rather than as a purely external annotation added after scaffold construction.

Branch assignment and scaffold program composition

To identify broad divergence structure within the diagnosis-defined latent scaffold, samples were assigned to branch-like basins based on their location in latent space and associated scaffold-program composition. These basins were labeled using biologically interpretable state descriptors corresponding to HSC-like, progenitor-like, GMP-like and Mono/DC-like organization. Branch labels were used as coarse-grained disease-state identities for longitudinal analysis and were not intended to imply exact lineage ancestry. Branch assignments and scaffold coordinates are provided in Supplementary Data 2.

For each branch, we calculated mean standardized scaffold-program values and summarized branch-specific ecological composition. These summaries were used to determine whether malignant-state basins were embedded in distinct ecological backgrounds and whether they differed in programmatic identity. Branch-specific ecological context and scaffold program composition are shown in Fig. 5B,C and provided in Supplementary Data 2. For longitudinal analyses, DX→REL intervals were classified as branch-continuous when diagnosis and relapse samples occupied the same broad branch basin and branch-switching when relapse occupied a different basin. These branch-transition annotations supported downstream analyses of displacement, jump-sensitive behavior and branch-conditioned escape propensity, with interval-level summaries provided in Supplementary Data 3.

Longitudinal projection and trajectory reconstruction

Non-diagnosis samples from the discovery cohort were projected into the frozen diagnosis-anchored scaffold using the same feature definitions, scaling parameters and latent mapping learned during scaffold construction. This enabled direct comparison of DX, EOI/REM and REL samples within a

common reference frame. Patient-level trajectories were reconstructed by linking projected timepoints within each individual in temporal order. This frozen-space strategy was used to ensure that longitudinal movement reflected displacement relative to the diagnosis-defined baseline rather than changes introduced by refitting or co-embedding later timepoints.

Trajectory summaries were examined at both the individual and group levels. Individual trajectories were used to visualize heterogeneous patient courses through latent space, whereas group-level summaries were based on phase-specific point clouds, median locations and representative interval arrows. DX→REL intervals were further summarized by total displacement, block-wise malignant-state displacement and ecological or tumor-microenvironment displacement. These interval-level trajectory and displacement summaries are provided in Supplementary Data 3, with additional branch-transition and displacement diagnostics shown in Supplementary Fig. S5.

The same projection logic was used for external validation analyses. GSE235923 samples were projected into the frozen discovery scaffold for independent AML calibration, and GSE227122 samples were projected for strict cross-lineage transfer. External AML projection outputs and cross-lineage transfer coordinates are provided in Supplementary Data 5, with supporting cross-lineage transfer diagnostics shown in Supplementary Fig. S7. This projection framework allowed disease progression and transfer behavior to be interpreted as movement relative to a shared diagnosis baseline rather than as timepoint-specific or cohort-specific clustering behavior.

Compact therapy-aware OU–Lévy–branching formulation

The therapy-aware OU–Lévy–Branching model should be viewed as a conceptual statistical scaffold rather than a fully identifiable continuous-time stochastic process. Given sparse clinical sampling, the latent dynamics are summarized through interpretable quantities reflecting attraction, displacement, instability, branch switching, and jump-sensitive behavior.

We considered a unified therapy-aware ecological model in which within-basin stabilization, branch redirection and rare jump-like departures jointly shape leukemia trajectories. Let X_t denote the reduced leukemia disease-state vector, e_t the ecological context, b_t the branch or latent trajectory state, and τ_t the treatment or clinical phase. In compact form, the latent dynamics were represented as

$$dX_t = \Theta_{(b_t,e_t,\tau_t)}\left(\mu_{(b_t,e_t,\tau_t)} - X_t\right)dt + \Sigma^{1/2}_{(b_t,e_t,\tau_t)}dW_t + dJ_t$$

,

where $\mu_{(b_t,e_t,\tau_t)}$ and $\Theta_{(b_t,e_t,\tau_t)}$ define the phase-, branch- and ecology-specific local attractor and restoring strength, $\Sigma^{1/2}_{(b_t,e_t,\tau_t)}$ denotes the diffusion scale or covariance factor, dW_t represents continuous Gaussian fluctuation, and dJ_t represents jump-like punctuated departures. In biological terms, ecotype and branch context define the local disease-state basin, OU-like dynamics summarize constrained mean-reverting motion within that basin, therapy can shift the attractor or alter effective

constraint, and the jump component represents rare larger departures compatible with relapse-associated escape.

The continuous mean-reverting part of the model follows the classical OU lineage and later adaptive and stochastic-diffusion extensions,[23–29] whereas the jump-capable formulation is motivated by general Lévy-driven OU processes and jump-based analyses of continuous-trait evolution.[30,31] The branching component was implemented operationally through assignment of samples to broad scaffold basins and by classifying DX→REL intervals as branch-continuous or branch-switching according to whether projected diagnosis and relapse states occupied the same basin, consistent with branching-process views of cancer evolution.[32] More broadly, this formulation is compatible with systems-level notions of cancer attractors and state transitions.[33]

Because the available longitudinal datasets are sparsely sampled, we did not interpret this model as direct estimation of a full continuous-time Lévy process. Instead, the formulation served as an interpretable scaffold linking observable quantities—attractor displacement, effective restoring strength, effective instability, branch switching and jump-sensitive upper-tail behavior—to therapy-aware disease-state dynamics. Conceptually, the present model extends prior hybrid OU–branching and ecotype-aware pediatric leukemia analyses by incorporating therapy-aware modulation and an explicit jump-capable component within a multimodal scaffold.[34–36] A formal mathematical extension of this therapy-aware ecological OU–Lévy–Branching formulation is provided in Supplementary Note 4.

Dynamic summary metrics

Several interpretable summary metrics were derived from projected trajectories and local scaffold geometry. These quantities were used as descriptive summaries of latent disease-state dynamics rather than as direct biophysical parameter estimates.

Effective restoring strength (θ_{eff}) quantified the degree to which a sample remained constrained toward its local attractor. Larger values were interpreted as stronger restoration or confinement within the disease-state landscape.

Attractor displacement from the diagnosis baseline, or μ -shift, measured the distance between a sample's inferred local attractor and the diagnosis-defined reference position. This metric was used to distinguish diagnosis-anchored states, retained residual states and more displaced relapse-associated states.

Effective instability, or diffusion proxy (σ_{eff}), captured local stochastic spread or reduced constraint around the inferred attractor. This quantity was interpreted comparatively across phases and intervals rather than as an absolute biophysical diffusion coefficient.

Jump-sensitive score was used as an interval-level statistic summarizing punctuated departure relative to expected local continuous movement in the latent space. It was used primarily to compare branch-

continuous and branch-switching intervals and to rank branch-level escape propensity.

Relapse-escape score combined overall DX→REL displacement with an added branch-switching contribution to prioritize intervals most compatible with escape-like reorganization. This score was used descriptively to identify upper-tail departures rather than as a binary clinical classifier.

Together, these metrics provided the basis for the therapy-aware dynamic regimes used in the main analyses: response-like constrained states, residual persistent states and escape-prone relapse states.

Classification of branch-continuous and branch-switching transitions

For each DX→REL interval in the discovery cohort, diagnosis and relapse branch identities were assigned from the projected latent scaffold. Intervals were classified as branch-continuous when both timepoints occupied the same broad branch basin and as branch-switching when the relapse sample occupied a different basin from the diagnosis sample. These interval-level branch annotations are provided in Supplementary Data 3.

This classification was used to compare displacement distributions, jump-sensitive scores and branch-conditioned escape behavior. It was not intended to imply exact lineage conversion; rather, it distinguished trajectories that remained within the same coarse disease-state basin from those that moved into a qualitatively different region of the diagnosis-defined scaffold. Additional branch-transition and displacement diagnostics are shown in Supplementary Fig. S5, and upper-tail and heavy-tail model-support analyses are shown in Supplementary Fig. S6.

Upper-tail displacement and escape-subset definition

To identify the most extreme longitudinal departures, DX→REL displacement values were ranked across all evaluable intervals in the discovery cohort. An empirical upper-tail threshold was defined using the 90th percentile of the observed DX→REL displacement distribution. Intervals above this threshold were designated the upper-tail departure subset for descriptive analysis. Interval-level displacement metrics, relapse-escape scores and jump-candidate rankings are provided in Supplementary Data 3.

This non-parametric thresholding approach was chosen to isolate the most displaced cases without imposing a distributional assumption on the DX→REL displacement values. Because the number of longitudinal intervals was modest, upper-tail analyses were interpreted as prioritization and structure-finding procedures rather than definitive inferential classification. Additional upper-tail diagnostics, robust effect-size summaries and Gaussian-tail surprisal rankings are shown in Supplementary Fig. S6A–D.

External AML projection and calibration

To assess generalizability, samples from the independent pediatric AML cohort GSE235923 were projected into the frozen discovery scaffold using the same feature definitions, scaling parameters and latent mapping learned from diagnosis-phase discovery samples. No relearning or cohort-specific refitting of the latent manifold was performed. This provided a direct test of whether the learned disease-state geometry captured transferable structure rather than cohort-specific organization. Cross-cohort calibration of ecotype and malignant-state summaries is shown in Supplementary Fig. S5, and external AML projection outputs are provided in Supplementary Data 5.

External samples were analyzed using the same phase labels and dynamic summary metrics used in the discovery cohort. We specifically evaluated whether EOI/REM samples remained near the diagnosis-defined attractor on average, whether relapse samples were more displaced than EOI/REM states, and whether residual-like states remained diffusely organized rather than collapsing into a single tightly restored cluster. Fully longitudinal external cases were highlighted to visualize movement through the frozen scaffold. External AML dynamic-parameter summaries, attractor-displacement estimates and full-triad annotations are provided in Supplementary Data 5.

Conservative serial bulk validation

To test whether the major disease-state ordering inferred from multimodal single-cell data remained detectable in a more clinically deployable but lower-resolution setting, we analyzed the serial bulk RNA-seq cohort GSE163634 as a conservative validation. Because bulk measurements do not preserve the same single-cell multimodal geometry, this analysis was intentionally restricted to reduced, interpretable coordinates derived from the scaffold, particularly attractor displacement and restoring-strength summaries. Leukemia-versus-control distributions of transferable validation variables are shown in Supplementary Fig. S2, and serial bulk validation outputs are provided in Supplementary Data 6.

Bulk samples were positioned relative to the discovery-derived disease-state axes using the same diagnosis-referenced logic applied to the single-cell cohorts. The goal of this validation was not to replicate branch-level structure or fine-grained latent geometry, but to determine whether control, diagnosis leukemia and response states preserved the expected directional ordering under the scaffold. This was treated as a stringent translational test because successful recovery of disease-state directionality under reduced resolution would indicate that the central scaffold logic is not confined to the highest-resolution assay setting. Bulk-validation summaries, serial response comparisons and transferred state-variable outputs are provided in Supplementary Data 6.

Clinical dynamic map and representative state summaries

For translational interpretation, we constructed a reduced clinical dynamic map based on attractor displacement from the diagnosis baseline and effective restoring strength. This two-axis representation was used to visualize response-like constrained states, residual persistent states and more displaced escape-prone relapse-like states in a clinically interpretable format. Dynamic-parameter summaries and state-tier assignments are provided in Supplementary Data 4.

Representative samples were further summarized in a scorecard including scaffold-derived dynamic metrics, branch identity, qualitative state tier, and jump-sensitive or branch-switching behavior. These visual summaries were designed to illustrate how the latent scaffold can be rendered as a clinically readable disease-state representation suitable for response monitoring and triage-oriented interpretation. The scorecard inputs, state-tier labels and clinical dynamic map variables are provided in Supplementary Data 4, with bulk-validation and model-comparison outputs provided in Supplementary Data 6.

Statistics and reproducibility

Analyses were primarily descriptive and structure-oriented, with formal comparisons restricted to prespecified contrasts supporting phase-level or transition-class interpretations. Comparisons between independent groups were performed using two-sided Mann–Whitney U tests. Matched diagnosis-to-relapse comparisons were performed using two-sided Wilcoxon signed-rank tests. Statistical tests were performed only when each independent group contained at least two evaluable observations or, for paired analyses, when at least two complete pairs were available; smaller groups were displayed individually and identified as exploratory. The nominal significance threshold was $\alpha = 0.05$. Exact unadjusted *P* values and group sample sizes are reported in the corresponding figures and legends. No adjustment for multiple comparisons was applied because the formal tests were prespecified and limited to comparisons of branch-continuous versus branch-switching intervals, diagnosis versus relapse states, diagnosis versus end-of-induction/remission states in the external cohort, and adjacent clinically ordered groups in the serial bulk validation cohort. Cliff's delta was calculated as a non-parametric effect-size measure for independent-group comparisons. Because several analyses involved modest subgroup sizes, particularly EOI/REM, relapse and branch-conditioned subsets, interpretation emphasized effect direction, geometric ordering and consistency across discovery and validation cohorts rather than isolated significance testing.

Software and reproducibility

All preprocessing, multimodal integration, latent-state construction, trajectory analysis, statistical summaries and figure generation were performed using reproducible scripted workflows in Python. The computational workflow incorporated standard libraries for array computation, data handling, visualization and single-cell analysis, together with custom code for scaffold construction, frozen-space projection, branch analysis, dynamical summary estimation, upper-tail analysis and model comparison. Sample-level preprocessing summaries, scaffold variables, dynamic parameters, external projection outputs, bulk-validation summaries and model-comparison tables are provided in Supplementary Data 1–6.

Declarations

Data availability

All datasets analyzed in this study are publicly available through the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE235063, GSE235923, and GSE163634. GSE227122 was used only in

the supplementary cross-lineage transfer analysis shown in Supplementary Fig. S7; because it contained only one relapse sample, it was not included in the main relapse-dynamics analyses. Processed data used for analysis are available through the GEO series records and associated supplementary files provided by the original studies. For GSE235063, controlled-access raw data are available through the European Genome-phenome Archive under accession EGAS00001007323. Where applicable, access conditions for raw or protected data are governed by the originating repositories and the policies of the source studies.

Code availability

All code used for data processing, multimodal scaffold construction, longitudinal projection, dynamical summary estimation, statistical analysis, and figure generation is publicly available from the project GitHub repository at https://github.com/shkim9391/Ecotype_Levy_OU_Branching_Dynamics. A DOI-backed archived release corresponding to the submission version is available at Zenodo: https://doi.org/10.5281/zenodo.21753569.

Acknowledgements

This study received no funding.

Author contributions

S.-H.K. conceived the study, designed the analytical framework, curated and analyzed the data, developed the computational workflow, generated the figures, interpreted the results, and wrote the manuscript.

Competing interests

The author declares no competing interests.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Supplementary information

Supplementary Information accompanies this manuscript and includes Supplementary Notes 1–4 and Supplementary Figures S1–S7. Supplementary Data 1–6 are provided separately.

Abbreviations

DX, diagnosis; EOI, end of induction; REM, remission; REL, relapse; AML, acute myeloid leukemia; HSC, hematopoietic stem cell-like; GMP, granulocyte–monocyte progenitor-like; Mono/DC, monocyte/dendritic

cell-like; OU, Ornstein–Uhlenbeck

References

1. Nowell, P. C. The clonal evolution of tumor cell populations. Science, 194, 23–28 (1976). https://doi.org/10.1126/science.959840
2. Greaves, M. & Maley, C. C. Clonal evolution in cancer. Nature, 481, 306–313 (2012). https://doi.org/10.1038/nature10762
3. Beerenwinkel, N. et al. Cancer evolution: mathematical models and computational inference. Syst. Biol. 64, e1–e25 (2015). https://doi.org/10.1093/sysbio/syu081
4. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. 366, 883–892 (2012). https://doi.org/10.1056/nejmoa1113205
5. Ding, L. et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature, 481, 506–510 (2012). https://doi.org/10.1038/nature10738
6. Davis, A. et al. Tumor evolution: linear, branching, neutral or punctuated? Biochim. Biophys. Acta Rev. Cancer, 1867, 151–161 (2017). https://doi.org/10.1016/j.bbcan.2017.01.003
7. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. Cell, 153, 666–677 (2013). https://doi.org/10.1016/j.cell.2013.03.021
8. Sayyab, S. et al. Mutational patterns and clonal evolution from diagnosis to relapse in pediatric acute lymphoblastic leukemia. Sci. Rep. 11, 15988 (2021). https://doi.org/10.1038/s41598-021-95109-0
9. Bissell, M. J. & Hines, W. C. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat. Med. 17, 320–329 (2011). https://doi.org/10.1038/nm.2328
10. Gatenby, R. A. & Gillies, R. J. A microenvironmental model of carcinogenesis. Nat. Rev. Cancer, 8, 56–61 (2008). https://doi.org/10.1038/nrc2255
11. Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature, 501, 346–354 (2013). https://doi.org/10.1038/nature12626
12. Baryawno, N. et al. A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. Cell, 177, 1915–1932.e16 (2019). https://doi.org/10.1016/j.cell.2019.04.040
13. van Galen, P. et al. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. Cell, 176, 1265–1281.e24 (2019). https://doi.org/10.1016/j.cell.2019.01.031
14. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science, 352, 189–196 (2016). https://doi.org/10.1126/science.aad0501
15. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell, 171, 1611–1624.e24 (2017). https://doi.org/10.1016/j.cell.2017.10.044
16. Luca, B. A. et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell, 184, 5482–5496.e28 (2021). https://doi.org/10.1016/j.cell.2021.09.014

17. Packer, J. S. & Trapnell, C. Single-cell multi-omics: an engine for new quantitative models of gene regulation. Trends Genet. 34, 653–665 (2018). https://doi.org/10.1016/j.tig.2018.06.001
18. Wagner, A. et al. Revealing the vectors of cellular identity with single-cell genomics. Nat. Biotechnol. 34, 1145–1160 (2016). https://doi.org/10.1038/nbt.3711
19. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat. Biotechnol. 37, 1458–1465 (2019). https://doi.org/10.1038/s41587-019-0332-7
20. Lambo, S. et al. A longitudinal single-cell atlas of treatment response in pediatric AML. Cancer Cell, 41, 2117–2135.e12 (2023). https://doi.org/10.1016/j.ccell.2023.10.008
21. Mumme, H. L. et al. Single-cell analysis reveals altered tumor microenvironments of relapse- and remission-associated pediatric acute myeloid leukemia. Nat. Commun. 14, 6209 (2023). https://doi.org/10.1038/s41467-023-41994-0
22. Mumme, H. L. et al. Identification of leukemia-enriched signature through the development of a comprehensive pediatric single-cell atlas. Nat. Commun. 16, 4114 (2025). https://doi.org/10.1038/s41467-025-59362-5
23. Uhlenbeck, G. E. & Ornstein, L. S. On the theory of the Brownian motion. Phys. Rev. 36, 823–841 (1930). https://doi.org/10.1103/PhysRev.36.823
24. Doob, J. L. The Brownian movement and stochastic equations. Ann. Math. 43, 351–369 (1942). https://doi.org/10.2307/1968873
25. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. Evolution, 51, 1341–1351 (1997). https://doi.org/10.1111/j.1558-5646.1997.tb01457.x
26. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164, 683–695 (2004). https://doi.org/10.1086/426002
27. Hansen, T. F., Pienaar, J. & Orzack, S. H. A comparative method for studying adaptation to a randomly evolving environment. Evolution, 62, 1965–1977 (2008). https://doi.org/10.1111/j.1558-5646.2008.00412.x
28. Beaulieu, J. M. et al. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution, 66, 2369–2383 (2012). https://doi.org/10.1111/j.1558-5646.2012.01619.x
29. Blomberg, S. P. et al. Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. Am. Nat. 195, 145–165 (2020). https://doi.org/10.1086/706339
30. Masuda, H. On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. Bernoulli, 10, 97–120 (2004). https://doi.org/10.3150/bj/1077544605
31. Landis, M. J. et al. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Syst. Biol. 62, 193–204 (2013). https://doi.org/10.1093/sysbio/sys086

32. Durrett, R. Branching Process Models of Cancer. (Springer, 2015). https://doi.org/10.1007/978-3-319-16065-8
33. Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin. Cell Dev. Biol. 20, 869–876 (2009). https://doi.org/10.1016/j.semcdb.2009.07.003
34. Kim S.-H. Modeling constrained tumor evolution through hybrid Ornstein–Uhlenbeck and branching dynamics. iScience 29, 116970 (2026). https://doi.org/10.1016/j.isci.2026.116970
35. Kim, S.-H. A hybrid Ornstein-Uhlenbeck-Branching framework unifies microbial and pediatric tumor evolution. Front. Oncol. 16, 1727973 (2026). https://doi.org/10.3389/fonc.2026.1727973
36. Kim, S.-H. Single-cell immune ecotypes shape microenvironment-modulated evolutionary dynamics in pediatric leukemia. Research Square, https://doi.org/10.21203/rs.3.rs-9012769/v1 (2026).

Figures

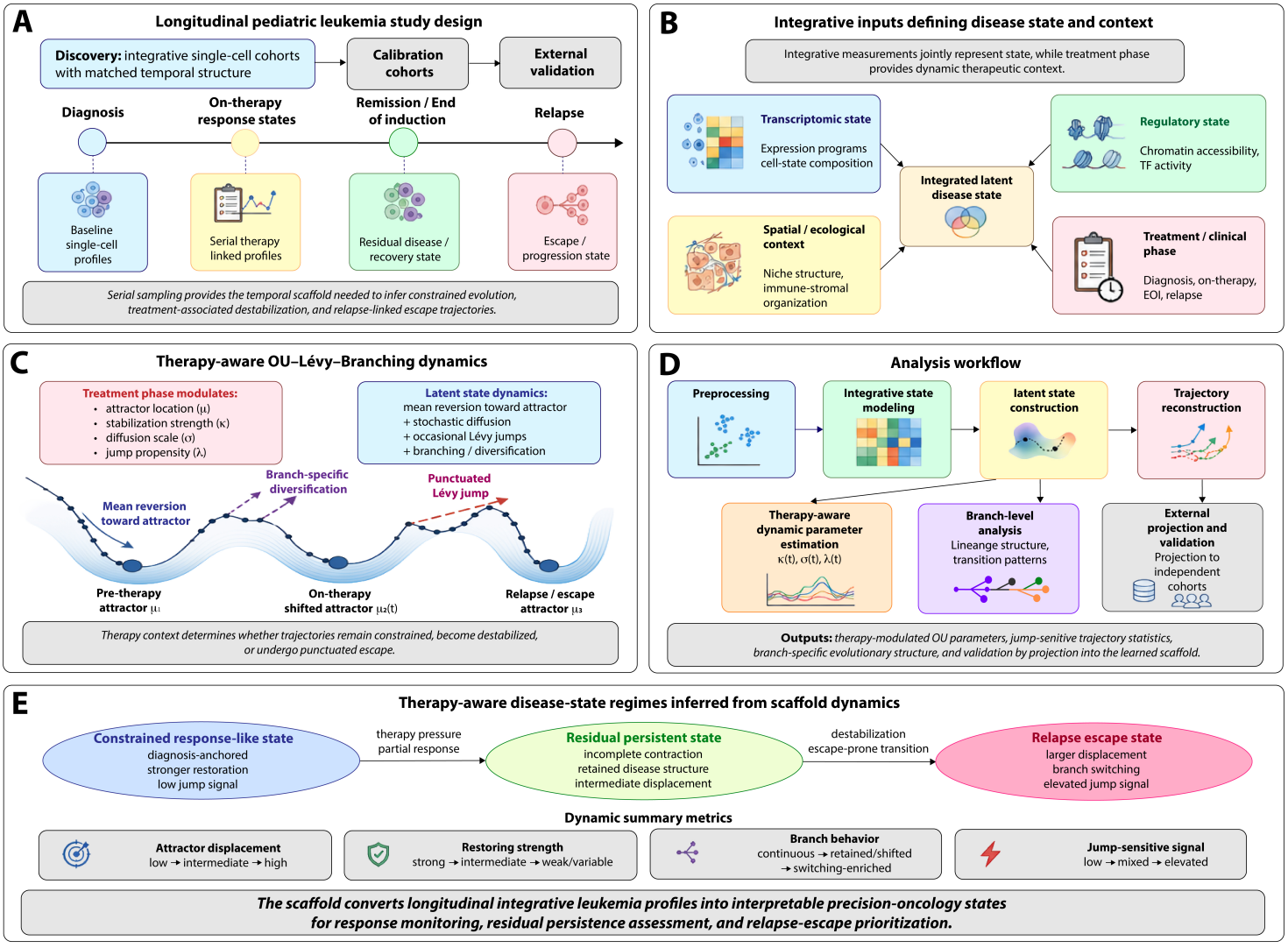
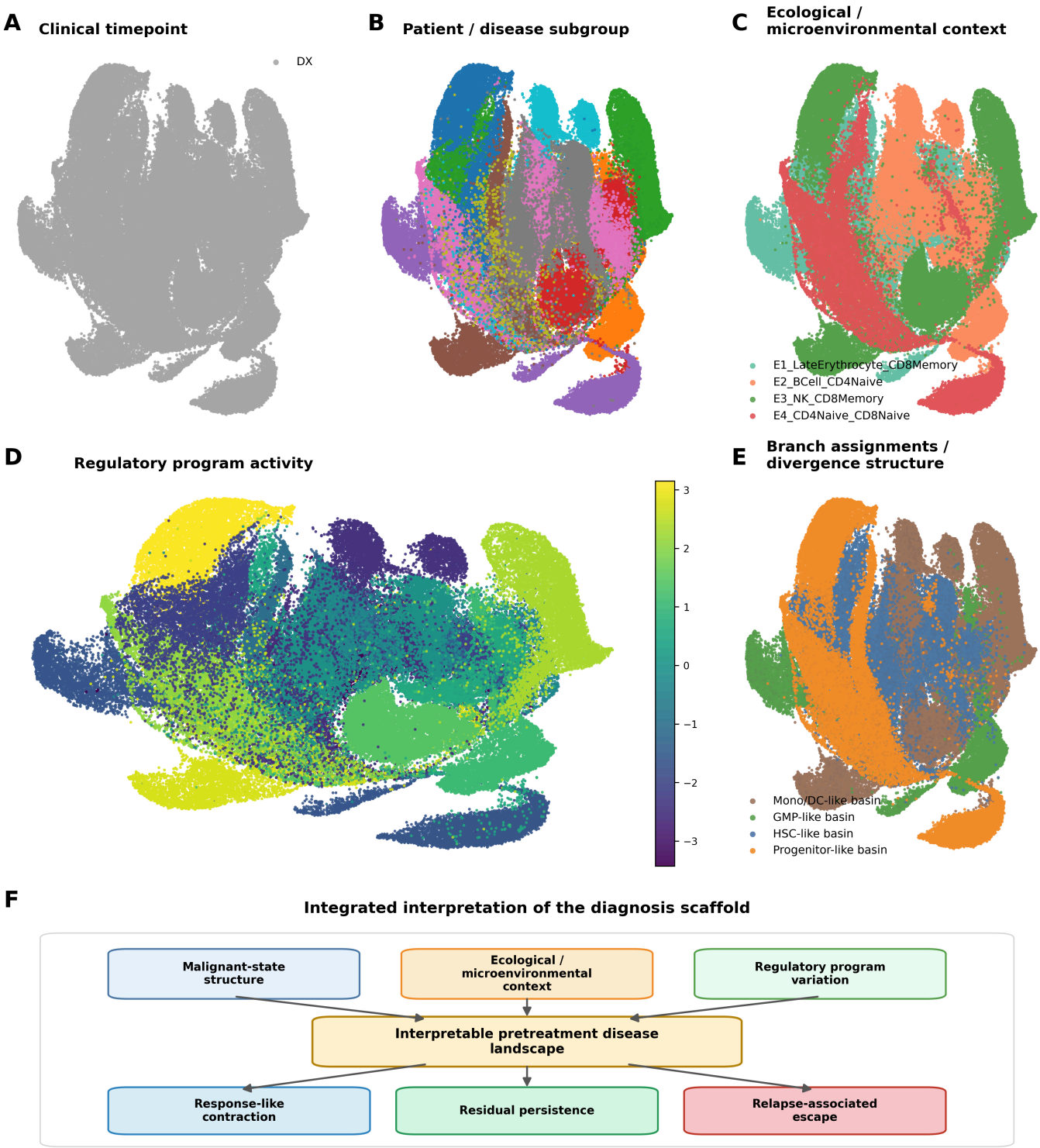


Figure 1

Study design and therapy-aware ecological OU–Lévy–branching framework for pediatric leukemia evolution. (a) Longitudinal study design spanning diagnosis, on-therapy response states, EOI/REM and relapse, with discovery, calibration and validation components. (b) Multimodal inputs used to define disease state, including transcriptomic, regulatory, ecological and clinical-phase information. (c) Therapy-aware OU–Lévy–branching model in which treatment phase can modulate attractor location, restoring strength, diffusion scale and jump propensity. (d) Analysis workflow from preprocessing and multimodal integration to trajectory reconstruction, dynamic-summary estimation, branch analysis and external validation. (e) Inferred disease-state regimes: response-like constrained states, residual persistent states and relapse escape states.

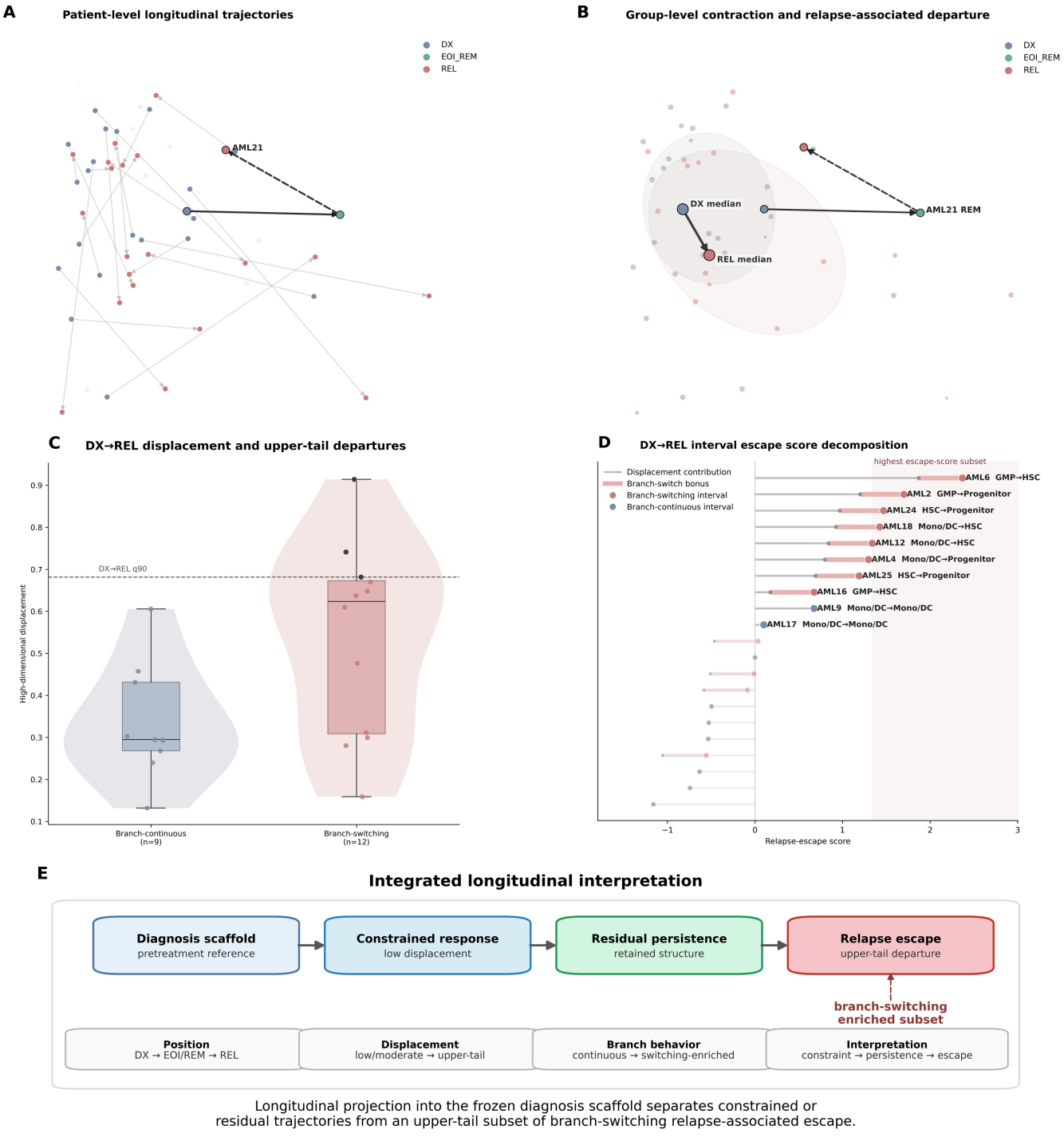


Malignant-state structure, ecological context and regulatory variation jointly define the pretreatment landscape used to quantify response-like contraction, residual persistence and relapse-associated escape.

Figure 2

Diagnosis-anchored scaffold captures malignant, ecological and regulatory disease-state organization. (a) Diagnosis-anchored scaffold used as the baseline reference for longitudinal projection. (b) Patient and disease-subgroup organization within the scaffold. (c) Ecological or microenvironmental contexts mapped onto the same latent space. (d) Regulatory program activity across the scaffold. (e) Branch assignments resolving HSC-like, progenitor-like, GMP-like and Mono/DC-like disease-state basins. (f)

Integrated interpretation: malignant-state structure, ecological context and regulatory variation jointly define the pretreatment landscape used to quantify response-like contraction, residual persistence and relapse-associated escape.

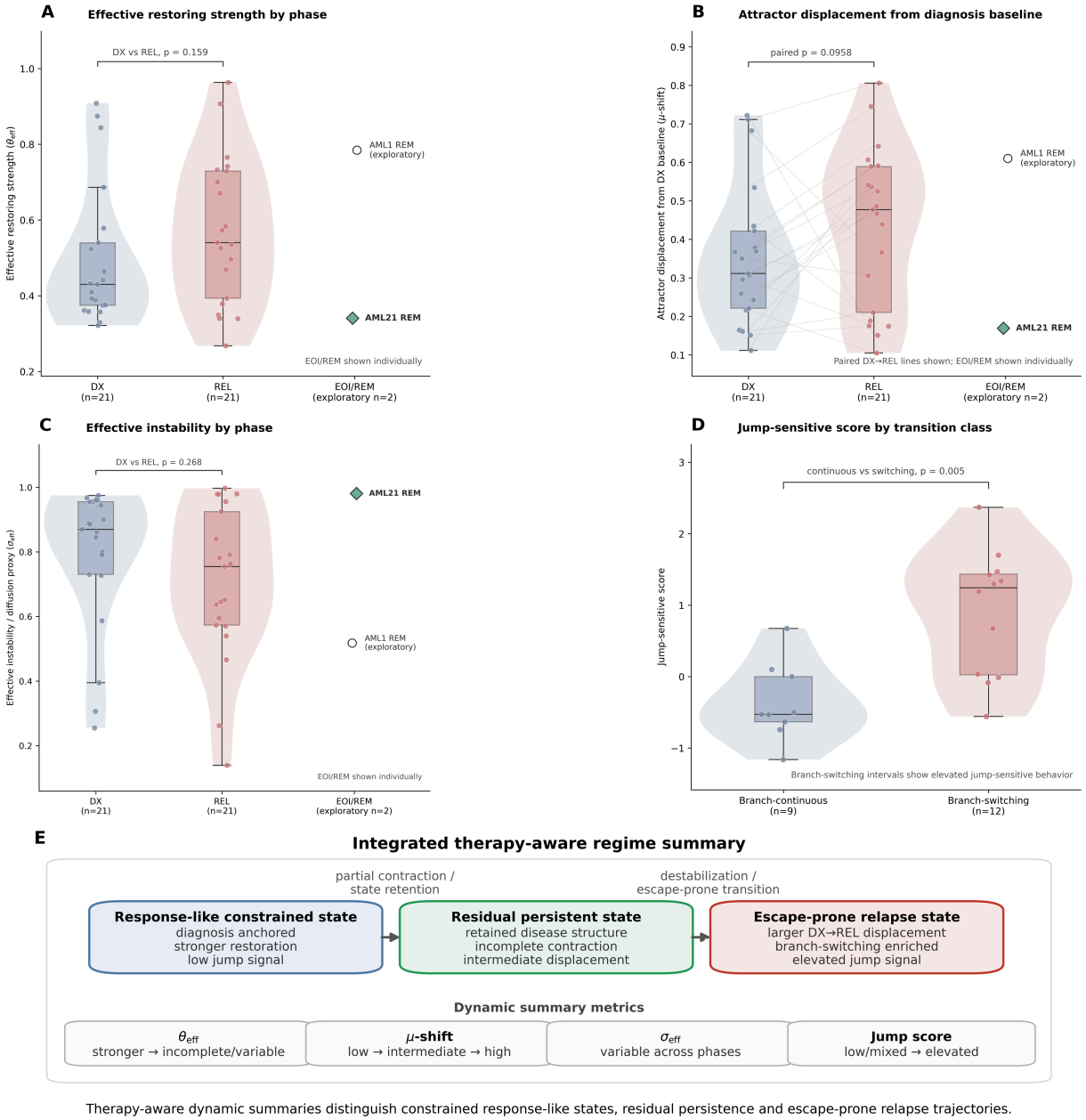


Longitudinal projection into the frozen diagnosis scaffold separates constrained or residual trajectories from an upper-tail subset of branch-switching relapse-associated escape.

Figure 3

Longitudinal projection identifies constrained response and relapse-associated departure. (a) Patient-level longitudinal trajectories projected into the frozen diagnosis scaffold. (b) Group-level organization of DX, EOI/REM and REL states in projected space. (c) High-dimensional DX→REL displacement for branch-

continuous (n=9) and branch-switching (n=12) intervals. Violin envelopes show kernel-density estimates; boxes indicate the interquartile range, center lines indicate medians, and whiskers extend to the most extreme observations within 1.5 times the interquartile range. Individual points represent matched DX→REL intervals. The dashed horizontal line indicates the empirical 90th percentile of the complete DX→REL displacement distribution, and dark points identify intervals at or above this threshold. Groups were compared using a two-sided Mann–Whitney U test; the displayed *P* value is unadjusted. (d) Relapse-escape score decomposition ranking intervals by displacement and branch-switching contribution. (e) Integrated interpretation showing constrained or residual trajectories together with an upper-tail subset of branch-switching relapse-associated escape.



Therapy-aware dynamic summaries distinguish constrained response-like states, residual persistence and escape-prone relapse trajectories.

Figure 4

Therapy-aware dynamical parameters separate response-like, residual persistent and escape-prone disease regimes. (a) Effective restoring strength, θ_eff, in main-analysis diagnosis (DX; n=21) and relapse (REL; n=21) samples. EOI/REM samples (n=2) are displayed individually and were not included in the pooled comparison. (b) Attractor displacement from the diagnosis baseline, or μ-shift, in DX (n=21) and REL (n=21) samples, with lines connecting complete patient-matched DX→REL pairs. The displayed comparison was performed using a two-sided Wilcoxon signed-rank test on complete pairs; EOI/REM samples (n=2) are displayed individually and were not included in the test. (c) Effective instability or diffusion proxy, σ_eff, in DX (n=21) and REL (n=21) samples; EOI/REM samples (n=2) are displayed individually. DX and REL distributions in panels A and C were compared using two-sided Mann–Whitney U tests. (d) Jump-sensitive scores for branch-continuous (n=9) and branch-switching (n=12) DX→REL intervals, compared using a two-sided Mann–Whitney U test. In panels A–D, violin envelopes show kernel-density estimates; boxes indicate the interquartile range, center lines indicate medians, whiskers extend to the most extreme observations within 1.5 times the interquartile range, and points represent individual samples or intervals. All displayed P values are exact and unadjusted. (e) Integrated therapy-aware regime summary showing response-like constrained, residual-persistent and escape-prone relapse states.

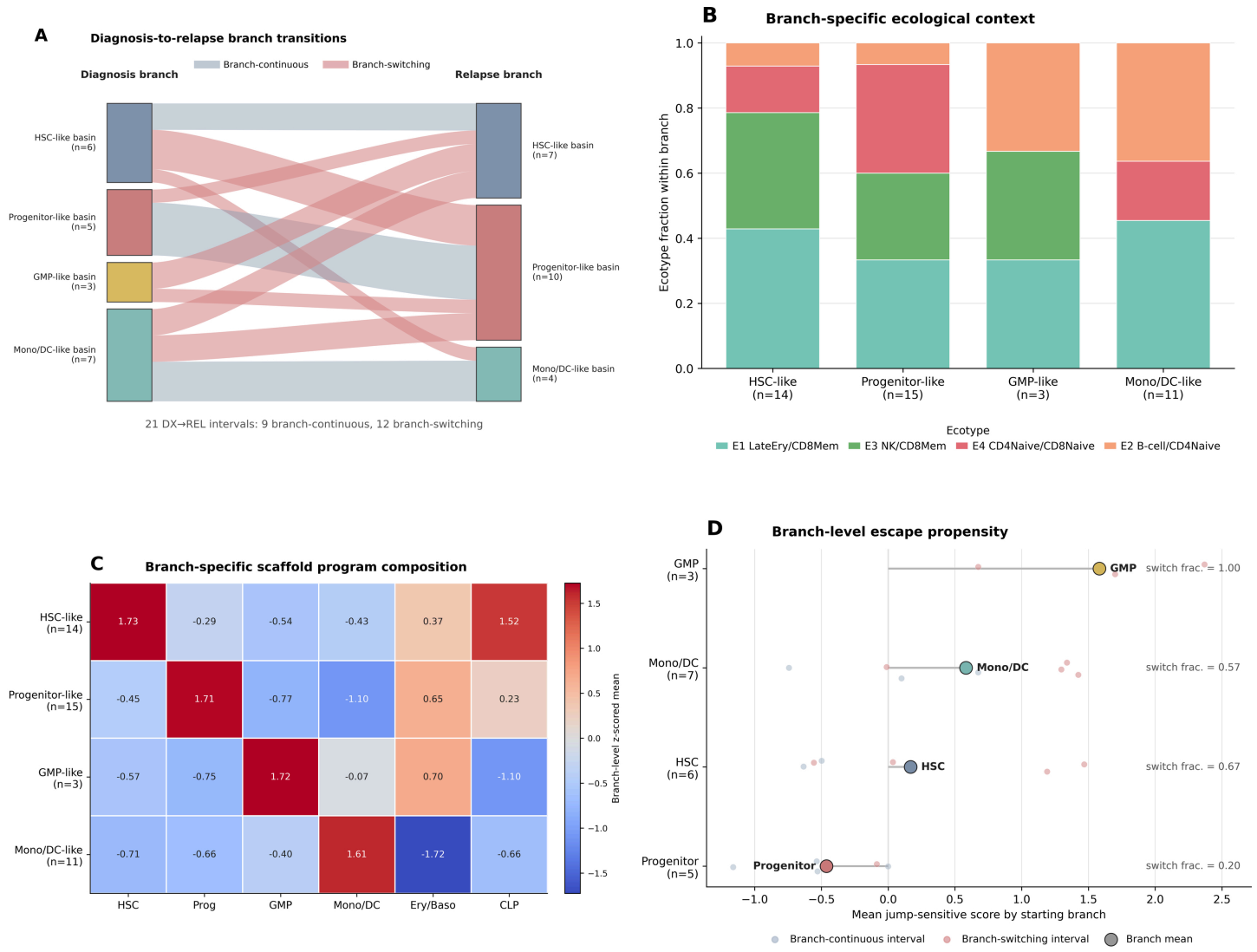


Figure 5

Branch structure and ecological context define non-equivalent routes to relapse-associated escape. (a) DX→REL branch transitions across 21 matched intervals, distinguishing branch-continuous and branch-switching routes. (b) Branch-specific ecological context shown as ecotype fractions within each branch. (c) Branch-level scaffold program composition shown as z-scored malignant-state and auxiliary program means. (d) Branch-level escape propensity, with individual interval jump-sensitive scores, branch means and branch-switching fractions.

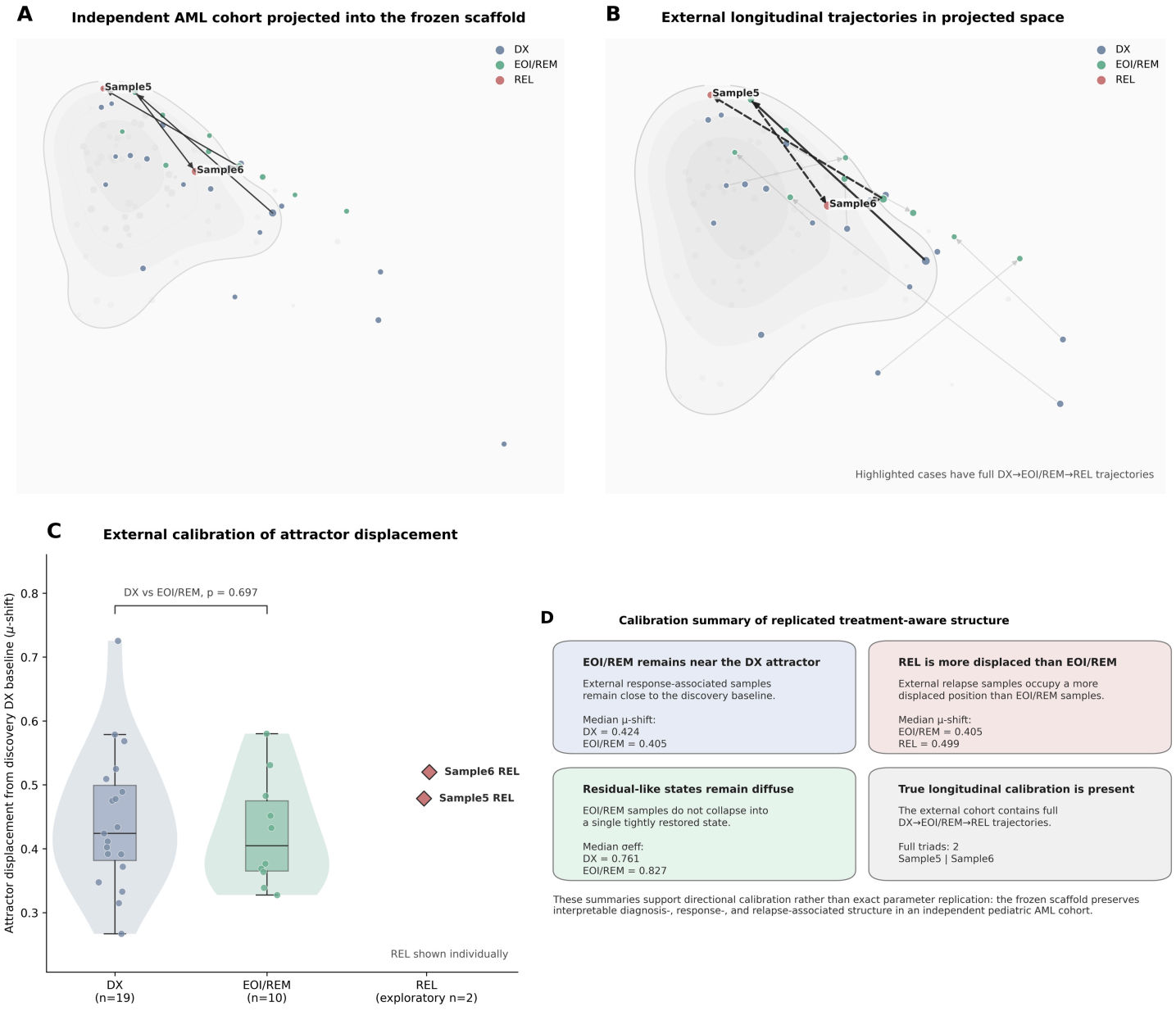


Figure 6

External AML projection reproduces treatment-aware organization in an independent cohort. (a) Independent pediatric AML samples projected into the frozen discovery scaffold without relearning the latent space. (b) External longitudinal trajectories in projected space, with complete DX→EOI/REM→REL trajectories highlighted. (c) Attractor displacement from the discovery DX baseline in external DX (n=19) and EOI/REM (n=10) samples. Sparse REL samples (n=2) are displayed individually and were not included in formal testing. Violin envelopes show kernel-density estimates; boxes indicate the interquartile range, center lines indicate medians, whiskers extend to the most extreme observations within 1.5 times the interquartile range, and points represent individual samples. DX and EOI/REM groups were compared using a two-sided Mann–Whitney U test; the displayed P value is unadjusted. (d) Summary of treatment-aware structure reproduced in the external cohort.

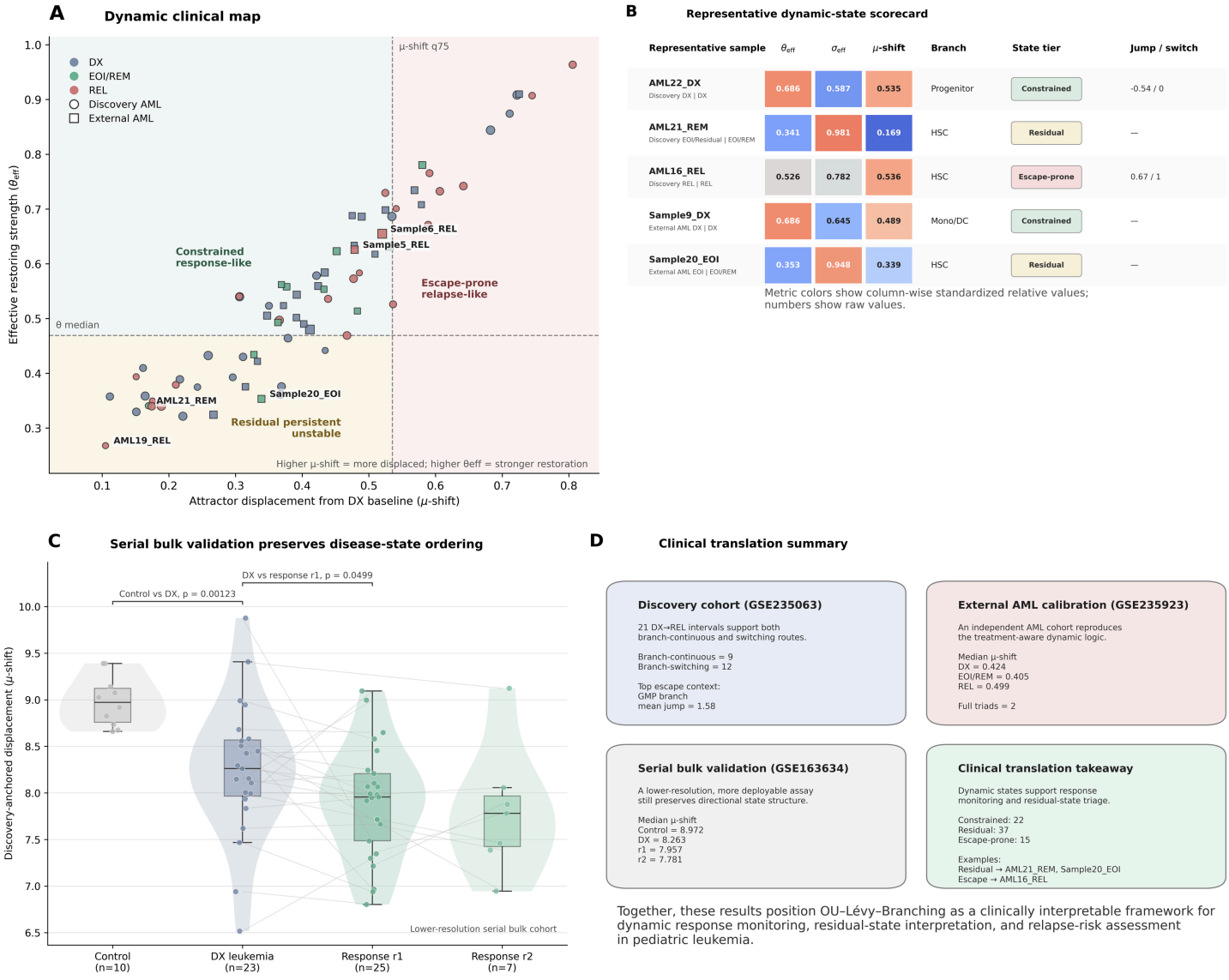


Figure 7

Clinical translation and conservative serial bulk validation of scaffold-derived disease-state structure. (a) Dynamic clinical map defined by attractor displacement from the diagnosis baseline and effective restoring strength. (b) Representative dynamic-state scorecard summarizing scaffold-derived metrics, branch identity, state tier and jump-sensitive or branch-switching behavior. (c) Discovery-anchored displacement in control (n=10), diagnosis (DX)-leukemia (n=23), response-r1 (n=25) and response-r2 (n=7) samples from the serial bulk RNA-seq cohort. Lines connect serial leukemia samples from the same patient where multiple stages were available. Violin envelopes show kernel-density estimates; boxes indicate the interquartile range, center lines indicate medians, whiskers extend to the most extreme observations within 1.5 times the interquartile range, and points represent individual samples. Prespecified independent-group comparisons were performed using two-sided Mann–Whitney U tests. The displayed comparisons are control versus diagnosis leukemia and diagnosis leukemia versus

response r1; displayed P values are unadjusted. The response-r1 versus response-r2 comparison was calculated and reported in Supplementary Data but was not displayed in the panel. (d) Clinical translation summary across discovery, external AML calibration and serial bulk validation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TherapyAwareEcologicalDynamicsPediatricLeukemiaSupplInformation.pdf
- SupplementaryData1.xlsx
- SupplementaryData2.xlsx
- SupplementaryData3.xlsx
- SupplementaryData4.xlsx
- SupplementaryData5.xlsx
- SupplementaryData6.xlsx
- SupplementaryDataGuide.xlsx