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.

Single-cell immune ecotypes shape microenvironment-modulated evolutionary dynamics in pediatric leukemia

Seung-Hwan Kim

seung-hwan.kim@fisher.edu

Department of Biology, Fisher College, Boston, MA 02116, USA; Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA https://orcid.org/0000-0002-9397-3324

Research Article

Keywords: Tumor microenvironment (TME), Single-cell RNA-seq, Pediatric leukemia, Stochastic evolutionary model, Ornstein–Uhlenbeck process, Clonal evolution/branching process, Bayesian inference, Immune ecotypes

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9012769/v1

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

Additional Declarations: The authors declare no competing interests.

Single-cell immune ecotypes shape microenvironment-modulated evolutionary dynamics in pediatric leukemia

Seung-Hwan Kim[1,2*]

[1]Department of Biology, Fisher College, Boston, MA 02116, USA
[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

*To whom correspondence should be addressed: seung-hwan.kim@fisher.edu

Keywords: Tumor microenvironment (TME), Single-cell RNA-seq, Pediatric leukemia, Stochastic evolutionary model, Ornstein–Uhlenbeck process, Clonal evolution/branching process, Bayesian inference, Immune ecotypes

Abstract

Tumor evolution is shaped not only by intrinsic genetic alterations but also by selective pressures imposed by the tumor microenvironment (TME). Although single-cell profiling increasingly resolves ecological heterogeneity, quantitative evolutionary models that explicitly incorporate microenvironmental structure remain limited. Here, we introduce a TME-modulated Ornstein–Uhlenbeck (OU) framework with ecotype-informed hierarchical structure to identify microenvironment-associated evolutionary regimes in pediatric leukemia. Using single-cell transcriptomes from the Single-cell Pediatric Cancer Atlas (ScPCA), we quantify patient-level TME composition and define reproducible immune ecotypes—recurring microenvironmental states that represent discrete ecological contexts. We fit an ecotype-informed OU model in which patient-specific trajectories along a pseudotime-derived transcriptional differentiation score evolve under stabilizing selection and stochastic diffusion, while ecotype membership modulates evolutionary parameters through hierarchical priors enabling partial pooling across patients with similar TMEs. Bayesian inference yields uncertainty-aware estimates of evolutionary optima and stabilizing selection strength at both patient and ecotype levels, and posterior predictive checks support model adequacy. Comparative analyses uncover distinct ecotype-specific regimes: immune-enriched microenvironments exhibit stronger stabilizing selection and more constrained dynamics, whereas stromal/other-enriched contexts show weaker stabilizing selection and greater stochastic drift. Together, these findings link single-cell–defined microenvironmental states to evolutionary dynamics and provide an interpretable, uncertainty-aware framework for comparing tumor evolution across patients and cohorts.

1. Introduction

Cancer evolution reflects an interplay between intrinsic genetic variation and extrinsic selective pressures imposed by the tumor microenvironment (TME) (Nowell, 1976; Gatenby and Gillies, 2008; Huang, 2009; Bissell and Hines, 2011; Greaves and Maley, 2012; Junttila and de Sauvage, 2013). While clonal diversification and stochastic drift are fundamental drivers of tumor progression, microenvironmental context can constrain phenotypic trajectories, shape context-dependent optima, and modulate the tempo of evolutionary change (Hansen, 1997; Butler and King, 2004; Gillies et al., 2012). Advances in single-cell profiling now resolve tumor ecosystems at unprecedented resolution, yet quantitative frameworks that integrate microenvironmental structure directly into evolutionary models remain limited (Tirosh et al., 2016; Puram et al., 2017; Wagner et al., 2016; Packer and Trapnell, 2018; Kimmel et al., 2021). Here, we use 'immune ecotypes' to denote recurring patient-level microenvironment composition states derived from single-cell cell-type fractions.

Most existing models of tumor evolution emphasize intrinsic dynamics—branching processes, random walks, or Markovian state transitions—and typically treat environmental effects implicitly or as unstructured noise (Nowell, 1976; Michor et al., 2005; McFarland et al., 2013; Beerenwinkel et al., 2015; Durrett, 2015). These approaches capture clonal diversification but often lack an explicit representation of stabilizing selection toward context-dependent phenotypic states. In contrast, continuous-time processes such as the Ornstein–Uhlenbeck (OU) process provide a natural description of evolution under stabilizing selection and are widely used in evolutionary biology to model adaptation toward ecological optima (Hansen, 1997; Butler and King, 2004; Hansen et al., 2008; Uyeda et al., 2011; Beaulieu et al., 2012). However, OU models alone do not represent lineage branching and clonal expansion—central features of cancer evolution—nor do they specify how discrete ecological contexts should modulate evolutionary parameters (Nowell, 1976; Michor et al., 2005; Beerenwinkel et al., 2015).

Recent work has begun to combine OU dynamics with branching structures to jointly model stabilizing forces and clonal diversification (Kim, 2025, 2026). A remaining challenge is how to incorporate structured ecological information—such as immune composition or stromal context—into such models in a principled, scalable manner (Gatenby and Gillies, 2008; Aktipis et al., 2013; Kather et al., 2018; West et al., 2018). In particular, it remains unclear whether discrete TME states impose reproducible constraints on evolutionary dynamics across patients, or whether distinct ecological contexts correspond to consistent evolutionary regimes (Gatenby et al., 2010; Marusyk et al., 2012; Beerenwinkel et al., 2015; Kather et al., 2018).

Pediatric leukemias provide a compelling setting to address these questions. Despite relatively low mutational burdens, pediatric leukemias display substantial phenotypic heterogeneity and variable clinical trajectories, consistent with an important role for non-genetic influences (Aifantis et al., 2008; Andersson et al., 2015; Gröbner et al., 2018; Ma et al., 2018). Single-cell studies further reveal marked inter-patient variation in immune and stromal composition, suggesting that leukemic populations occupy distinct microenvironmental niches (van Galen et al., 2019; Granja et al., 2019; Baryawno et al., 2019; Barkley et al., 2022). These observations motivate evolutionary models that explicitly condition on microenvironmental structure rather than treating it as residual variation.

Here, we introduce a TME-modulated OU modeling framework that links tumor microenvironmental ecotypes with evolutionary dynamics in pediatric leukemia. Using single-cell transcriptomic data from the Single-cell Pediatric Cancer Atlas (ScPCA), we derive patient-level TME composition profiles and identify reproducible immune ecotypes representing discrete ecological states. We then fit an ecotype-informed OU model in which latent evolutionary trajectories evolve under stabilizing selection and stochastic diffusion, while ecotype membership enters through hierarchical priors that modulate OU parameters and enable partial pooling across patients sharing similar microenvironmental contexts. The term "OU–Branching" reflects a broader motivation to embed OU dynamics within branching lineage structures; however, in this study, we focus on the ecotype-modulated OU component and assess adequacy via posterior predictive checks.

This framework quantifies how evolutionary drift and stabilizing selection vary across ecological states, while stochastic volatility is governed by a shared global diffusion scale. Applied to pediatric leukemia, it identifies ecotype-associated evolutionary regimes: immune-enriched microenvironments exhibit stronger stabilizing selection and more constrained dynamics, whereas stromal- or other-enriched contexts show weaker stabilizing selection and greater drift. By integrating single-cell–derived TME structure with uncertainty-aware OU modeling, our work provides a generalizable computational approach for studying how ecological context shapes cancer evolution.

2. Methods

2.1 Data sources and patient cohorts

All patient-level single-cell transcriptomic profiles and harmonized metadata were obtained from the Alex's Lemonade Stand Foundation Childhood Cancer Data Lab Single-cell Pediatric Cancer Atlas (ScPCA) Portal. We analyzed two publicly available pediatric leukemia projects hosted in ScPCA.

Project 1: Single-cell sequencing of acute lymphoblastic leukemia (ScPCA: SCPCP000008). This project contains 104 samples profiled with 10x Genomics 5′ (v2) and 3′ (v3.1) gene expression assays, with CITE-seq available for a subset. Diagnoses include B-cell acute lymphoblastic leukemia (B-ALL; n = 93), early T-cell precursor T-cell acute lymphoblastic leukemia (ETP-ALL; n = 5), mixed phenotype acute leukemia (n = 5), and T-cell acute lymphoblastic leukemia (T-ALL; n = 1). The portal provides associated sample- and participant-level annotations (e.g. disease ontology terms, risk group, karyotype, and additional covariates relevant to TME characterization). This project is linked to the cohort described in Iacobucci et al. (2025).

Project 2: Single-cell RNA sequencing of diverse pediatric leukemias (ScPCA: SCPCP000022). This project contains 25 samples profiled using 10x Genomics v2 single-cell RNA sequencing. Diagnoses

include B-ALL (n = 18), T-ALL (n = 4), and acute myeloid leukemia (AML; n = 3). Samples span diagnosis, remission, and recurrence/relapse. As described in the project record, samples were demultiplexed to generate sample-specific FASTQ files (via demuxlet using patient-specific variant calls) and processed with the ScPCA-nf workflow to produce standardized outputs and metadata for downstream analysis.

Across both projects, we retained only samples that satisfied downstream analysis requirements, including adequate cell yield, high-confidence annotations, and complete TME covariate information for immune ecotype inference and evolutionary modeling.

2.2 Trait construction (response variable)

The modeled trait is a scalar summary derived from ScPCA single-cell annotations for each sample (sample_id) at a given clinical timepoint (disease_timing). For each sample, we computed cell-type composition features as fractions of annotated cells. We included an "unknown" category for cells without a confident label (frac_unknown). For labeled compartments (T, myeloid, NK, B, stromal), we computed fractions conditional on known labels, i.e., among cells assigned to any known compartment (excluding unknown), yielding frac_<celltype>_given_known. Each fraction feature was then standardized across samples to improve comparability, producing z-scored variables (*_z; e.g., frac_T_given_known_z, frac_myeloid_given_known_z, frac_unknown_z).

In the OU–Branching analyses, we model the response Y as one chosen trait at a time (one model fit per trait), where each observation corresponds to a sample/timepoint for a participant (participant_id).

2.3 Tumor microenvironment (TME) feature extraction

TME feature construction

Tumor microenvironment (TME) covariates were derived from single-cell transcriptomes by aggregating ScPCA cell-type annotations to the participant level. When multiple samples/timepoints were available per participant, we aggregated sample-level compartment fractions to a single participant-level profile (one TME vector per participant) for ecotype inference. For each participant, we summarized the microenvironment using six compartment features: T cells, B cells, myeloid cells, natural killer (NK) cells, stromal cells, and unknown/other. Immune/stromal compartment fractions were computed conditional on confidently annotated ("known") cells (frac_*_given_known), while the unknown/other feature captured the proportion of cells without a confident compartment assignment (frac_unknown). These fraction features were then standardized across participants (z-scored; _z suffix) and used as ecological covariates in downstream analyses, including immune ecotype inference and TME-modulated OU–Branching modeling (Fig. 2A).

Correlation structure and dimensionality reduction

To quantify coupling among TME components, we computed pairwise Spearman rank correlations across participants (Fig. 2B). To visualize dominant axes of microenvironmental variation, we performed principal component analysis (PCA) on centered and scaled TME feature vectors (Fig. 2C and 3A). To capture nonlinear structure in the same feature space, we applied uniform manifold approximation and projection (UMAP) using Euclidean distance on the standardized TME vectors (Fig. 3B).

2.4 Identification of immune ecotypes

K-means clustering of ScPCA-derived TME states (immune ecotypes)

We applied k-means clustering to the participant-level standardized TME feature vectors (six z-scored compartment covariates; Section 2.3) to identify discrete immune ecotypes representing recurring microenvironmental states across the pediatric leukemia cohort. Ecotypes were assigned at the participant level. When multiple samples/timepoints were available for a participant, we aggregated sample-level TME composition features into a single participant-level profile prior to clustering, yielding one ecotype label per participant, which was used for all downstream analyses. The number of clusters was selected using an elbow analysis of within-cluster sum-of-squares, and we used $k = 4$ as a parsimonious representation of cohort-level TME structure (Fig. 2D). Ecotypes were annotated post hoc using their mean compartment profiles (Fig. 3C) and radar summaries (Fig. 3D), yielding four interpretable regimes: (1) T-cell–enriched, (2) mixed immune, (3) myeloid-polarized, and (4) stromal/other-enriched. Ecotype membership was subsequently incorporated as a higher-level grouping variable in downstream analyses using the TME-modulated OU–Branching framework.

2.5 Hybrid OU–Branching generative model

Model formulation

To quantify evolutionary pressure acting on leukemic populations, we model participant-level longitudinal trait dynamics using an Ornstein–Uhlenbeck (OU) process, motivated by OU–Branching generative formulations, and conditioned on ScPCA-derived tumor microenvironment (TME) state information (Fig. 1A). For a participant i, let $Y_{i,s}(t_{ij})$ denote the observed ScPCA-derived trait summary measured at sampling time t_{ij} for longitudinal series $\mathcal{S}$ (e.g., repeated trajectories per participant). Each observation corresponds to a sample/time point for the participant i. Ecotype membership is assigned once per participant and treated as fixed across that participant's samples, modulating hierarchical priors on μ_i and θ_i, while stochastic volatility is governed by a shared global diffusion scale. When repeated measurements are available, OU transitions between sampling times admit a closed-form Gaussian transition density. In particular, for $\Delta t = t_{ij} - t_{ij-1} > 0$,

$$Y_{i,s}(t_{ij}) \,\Big|\, Y_{i,s}(t_{ij-1}) \sim \mathcal{N}\left(\mu_i + \left(Y_{i,s}(t_{ij-1}) - \mu_i\right)e^{-\theta_i \Delta t}, \frac{\sigma_{proc}^2}{2\theta_i}\left(1 - e^{-2\theta_i \Delta t}\right)\right),$$

which we use for likelihood construction and posterior predictive simulation. Here, μ_i is the participant-specific evolutionary optimum (drift target), $\theta_i > 0$ controls the strength of mean reversion toward μ_i, and $\sigma_{proc} > 0$ is a global diffusion scale governing stochastic volatility.

Ecotype-informed hierarchical structure

Immune ecotypes inferred from ScPCA-derived TME composition vectors were incorporated as group-level structure to partially pool OU parameters across participants with similar microenvironmental context (Fig. 1C). If a participant i belongs to an ecotype $\kappa(i)$, and X_i denotes the vector of TME covariates, we model

$$\mu_i = \mu_0 + \alpha_{k(i)} + X_i^\top \beta,\ log\theta_i = \theta_0 + \gamma_{k(i)} + X_i^\top \eta,$$

with $\theta_i = exp(log\theta_i)$. Here, α_k and γ_k are ecotype-level random intercepts for the evolutionary optimum and selection strength, respectively, while β and η capture TME covariate effects. This hierarchy borrows statistical strength across participants within the same TME-defined regime while preserving participant-specific parameterization through ecotype and covariate effects. In the implemented model, stochastic volatility is governed by a shared σ_{proc} across participants.

Bayesian inference and posterior predictive evaluation

Posterior inference was performed using Hamiltonian Monte Carlo (No-U-Turn sampler) implemented in PyMC. We ran multiple independent chains with warm-up and monitored convergence using $\hat{R}$, effective sample size, and trace diagnostics for all primary parameters. Model adequacy was assessed using posterior predictive checks: replicated observations and/or trajectories were generated from the posterior predictive distribution and compared with observed ScPCA-derived summaries, with predictive bands and empirical coverage reported (Fig. 1D). Posterior samples of $(\mu_i, \theta_i, \sigma_{proc})$ were subsequently used for ecotype-stratified comparisons and downstream analyses.

2.6 Baseline and ablation models for predictive evaluation

To quantify the predictive contribution of OU mean-reverting dynamics and ecotype/TME-informed structure, we compared the full ecotype-informed OU model (TME-OU) against targeted baselines and ablations (Fig. 5). As a reference, the OU-only baseline retained the same observation model and OU latent-trajectory specification but removed ecotype/TME modulation (i.e., no ecotype-dependent priors on OU parameters). To test whether ecotype structure carries a signal beyond model flexibility, we fit a shuffled-TME control in which participant ecotype labels were

randomly permuted prior to inference, preserving marginal ecotype frequencies while breaking participant–TME coupling. We additionally evaluated two non-OU alternatives: a random-walk latent model (diffusive latent dynamics without mean reversion) and a static ecotype-only model that omits temporal dynamics and models observations using ecotype membership alone.

Predictive performance was evaluated using complementary criteria on the same observation set (Fig. 5). First, we computed PSIS-LOO expected log predictive density (ELPD) from pointwise log-likelihood contributions and reported ΔELPD relative to the OU-only baseline (higher is better; Fig. 5A). Uncertainty in ΔELPD was summarized using the standard error of pointwise LOO differences. Second, we assessed point prediction accuracy using RMSE between observed values and posterior predictive means (lower is better; Fig. 5B). Third, we evaluated uncertainty calibration by computing the empirical coverage of nominal 95% posterior predictive intervals (target ≈ 0.95; Fig. 5C). We monitored PSIS diagnostic stability via Pareto-k values and interpreted LOO-based comparisons cautiously when influential observations yielded elevated k. All models used identical preprocessing and identical posterior predictive evaluation code to ensure a controlled comparison.

2.7 Visualization of OU dynamics by ScPCA ecotype

Marginal posterior distributions

We compared ecotype-stratified marginal posteriors for the OU drift mean μ and stabilizing selection strength θ using violin plots (Fig. 4A–D). For each ecotype, violins summarize the posterior draws of the corresponding parameter, and participant-level posterior means were overlaid as points to visualize within-ecotype heterogeneity. Points were color-coded by clinical diagnosis metadata to highlight diagnosis-associated patterns.

Joint parameter space

To examine ecotype-specific evolutionary regimes in a multivariate context, we plotted each participant's posterior mean μ against $log_{10}\theta$ (Fig. 4E). Participants with extreme posterior estimates relative to their ecotype distribution were annotated to facilitate interpretation. Ecotype centroids were computed as the mean of participant posterior means within each ecotype and visualized with ±1 standard deviation in both dimensions, summarizing characteristic evolutionary landscapes associated with each TME state.

Flagship case study and robustness analyses

For the flagship analysis (Fig. 6), we summarized ecotype prevalence and TME composition by reporting the fraction of participants assigned to each ecotype and the ecotype-wise mean of standardized (z-scored) TME covariates (Fig. 6A). To connect inferred parameters to observed dynamics, we visualized representative participant trajectories by overlaying observed time-ordered summaries with the posterior predictive median and 95% posterior predictive intervals

(Fig. 6D). To assess robustness to ecotype granularity, we repeated unsupervised clustering over a range of k and evaluated how cluster-level $log_{10}\theta$ differed across alternative partitions (Fig. 7).

2.8 Model interpretation

Integrating ScPCA-derived immune ecotypes with ecotype-modulated OU modeling indicates that tumor microenvironmental heterogeneity is associated with substantial variation in inferred evolutionary dynamics across pediatric leukemias. In particular, immune ecotypes exhibited distinct evolutionary signatures, including (i) shifts in inferred evolutionary optima (μ), (ii) differences in the strength of stabilizing selection (θ), and (iii) a shared global diffusion scale (σ_{proc}) governing stochastic volatility. Collectively, these results support an interpretable link between ScPCA-resolved microenvironmental context and participant-specific evolutionary regimes, providing an ecological perspective on heterogeneity in disease progression, relapse trajectories, and variability in treatment response.

3. Results

3.1 ScPCA single-cell leukemia datasets reveal extensive microenvironmental diversity across patients

We first characterized tumor microenvironment (TME) diversity across pediatric leukemias using two publicly available single-cell datasets from the Single-cell Pediatric Cancer Atlas (ScPCA) Portal, comprising 104 ALL/MPAL samples and 25 additional leukemia samples spanning B-ALL, T-ALL, and AML. Harmonized cell-type annotations and clinical metadata produced by the ScPCA-nf pipeline enabled consistent aggregation of major compartments—including T cells, B cells, myeloid cells, NK cells, stromal populations, and unknown/other—across patients.

Aggregating these annotations into patient-level TME compartment features revealed pronounced ecological heterogeneity across the cohort (Fig. 2A). Lymphoid infiltration varied substantially between patients, while myeloid and stromal compartments showed a strong positive association (Fig. 2B), consistent with coordinated microenvironmental remodeling within leukemic niches.

Dimensionality reduction further indicated that patients are distributed along interpretable ecological gradients. PCA separated samples along axes corresponding to lymphoid-enriched versus stromal/myeloid-dominated microenvironments, with diagnostic subgroups occupying distinct but partially overlapping regions of TME space (Fig. 2C). A complementary UMAP embedding showed that TME variation is organized into discrete regions rather than a single continuous continuum (Fig. 3B), supporting the presence of recurrent, structured microenvironmental states across pediatric leukemias.

3.2 Four reproducible ScPCA-derived immune ecotypes define the ecological landscape of pediatric leukemias

Using six-dimensional ScPCA-derived TME compartment profiles, we applied *k*-means clustering and identified four reproducible immune ecotypes (Fig. 2D and Fig. 3). These ecotypes represent recurring microenvironmental states observed across multiple leukemia diagnoses, including B-ALL, T-ALL, and AML.

The four immune ecotypes exhibited distinct compositional signatures (Fig. 3C,D):

- Ecotype 1: T-cell–enriched, lymphoid-dominant microenvironments
- Ecotype 2: mixed immune composition
- Ecotype 3: myeloid-polarized, inflammatory microenvironments
- Ecotype 4: stromal- and unknown-enriched microenvironments

Radar and stacked-bar summaries revealed clear differences in mean compartment composition across ecotypes (Fig. 3C,D). Importantly, ecotype structure was not reducible to diagnostic labels alone, as multiple diagnoses occupied overlapping regions of ecotype-defined TME space (Fig. 3A,B). These ecotypes were subsequently used as group-level ecological structure to modulate priors in the ecotype-informed OU modeling framework.

3.3 Ecotype-informed hybrid OU–Branching modeling captures patient-specific evolutionary dynamics across ScPCA leukemias

We next fit the ecotype-informed OU model to each ScPCA patient to infer evolutionary pressures acting on leukemic populations. The framework models latent trait trajectories using continuous-time Ornstein–Uhlenbeck (OU) dynamics—capturing drift toward an inferred optimum under stabilizing selection with stochastic diffusion—while incorporating immune ecotype labels as a hierarchical structure to encode tumor microenvironmental context through ecotype-modulated priors (Fig. 1A,C). Although motivated by OU–Branching generative formulations, the analyses presented here focus on the ecotype-modulated OU component.

Posterior predictive checks showed close agreement between observed ScPCA-derived trait summaries and trajectories generated from the posterior predictive distribution (Fig. 1D), supporting model adequacy for the single-cell cohorts. Predictive performance and uncertainty calibration were further assessed using baseline and ablation comparisons (Section 3.6; Fig. 5), and model interpretability is illustrated in a flagship case study (Section 3.7; Fig. 6).

For each patient, we inferred posterior distributions for drift/optimum μ_i and stabilizing selection strength θ_i, along with a single shared diffusion scale (σ_{proc}) that governs stochastic volatility across the cohort. Together, these parameters quantify how strongly patient-specific dynamics are constrained toward ecotype-associated optima and the extent to which variation arises from stochastic fluctuations.

3.4 ScPCA immune ecotypes exhibit distinct evolutionary regimes

Posterior estimates of the OU drift mean μ differed across immune ecotypes, indicating ecotype-associated shifts in inferred evolutionary optima (Fig. 4A,B). Ecotype 1 (T-cell–enriched microenvironments) showed higher μ values on average, whereas Ecotype 4 (stromal/unknown-enriched microenvironments) exhibited lower μ and greater between-patient dispersion, consistent with weaker constraint toward a shared optimum.

Stabilizing selection strength θ, visualized on a logarithmic scale, also varied systematically across ecotypes (Fig. 4C,D). Ecotype 1 displayed comparatively stronger stabilizing selection, while Ecotype 3 (myeloid-polarized microenvironments) showed substantial heterogeneity in θ, suggesting variable constraint across patients within this ecological context. Ecotype 4 tended toward lower θ, consistent with more drift-dominated dynamics.

Within ecotypes, diagnostic categories (B-ALL, T-ALL, and AML) occupied distinct but overlapping regions of posterior parameter space (Fig. 4E), indicating that both intrinsic disease context and extrinsic microenvironmental state contribute to patient-specific evolutionary regimes.

3.5 Joint OU parameter space reveals ecotype-specific evolutionary attractors derived from ScPCA data

Visualization of joint posterior parameter space—posterior mean drift μ versus $log_{10}\theta$—showed ecotype-associated structure, with patients clustering in regions that closely corresponded to ScPCA-derived immune ecotypes (Fig. 4E). These clusters summarize ecotype-specific evolutionary regimes characterized by distinct combinations of inferred optima and stabilizing constraint:

- Ecotype 1: higher μ and moderate-to-high θ, consistent with stronger stabilizing selection in lymphoid-rich microenvironments.
- Ecotype 2: intermediate μ and θ, consistent with an intermediate regime between lymphoid- and myeloid/stromal-dominated states.
- Ecotype 3: lower μ with broad variation in θ, indicating heterogeneous constraint across patients within myeloid-polarized contexts.
- Ecotype 4: low μ and low θ, consistent with weaker mean reversion and more drift-dominated dynamics in stromal/unknown-enriched niches.

Patients deviating markedly from their ecotype centroid in (μ, $log_{10}\theta$) space were annotated in Fig. 4E to highlight inter-patient variability in inferred evolutionary dynamics. Such deviations may reflect atypical ecological constraints or patient-specific microenvironmental shifts associated with disease state and treatment history, motivating follow-up analyses in larger longitudinal cohorts.

3.6 Ablation and baseline comparisons evaluate OU dynamics and ecotype-informed structure

We next evaluated the predictive behavior of the ecotype-informed TME-OU model relative to OU dynamics alone using targeted baselines and ablations (Fig. 5). Using PSIS-LOO, we computed ΔELPD relative to the OU-only baseline (Fig. 5A; error bars denote the standard error of pointwise LOO differences). Models lacking OU mean reversion performed substantially worse: the random-walk latent model showed a marked reduction in predictive fit and higher RMSE, indicating that mean-reverting dynamics are important for capturing the observed trajectory structure. The static ecotype-only model also exhibited lower point prediction accuracy than OU-based models (Fig. 5B).

Among OU-based variants, predictive differences were small relative to uncertainty (Fig. 5A,B). The OU-only model achieved the highest PSIS-LOO in this evaluation, whereas the full TME-OU model was statistically indistinguishable from OU-only within ΔELPD uncertainty and showed comparable RMSE and near-nominal uncertainty calibration (Fig. 5B,C). The shuffled-TME control yielded similar predictive performance, suggesting that—in this dataset—the primary contribution of ecotype-informed structure is improved regularization and interpretability (ecotype-level regime inference and partial pooling) rather than consistent gains in out-of-sample predictive accuracy.

Uncertainty calibration further supported the OU-based formulation (Fig. 5C). OU-based models produced empirical coverage close to the nominal 95% target, whereas the static ecotype-only baseline showed over-coverage, consistent with overly conservative predictive intervals. Because PSIS diagnostics indicated a small number of influential observations (elevated Pareto-k), we interpret fine-grained differences among OU variants cautiously. Collectively, these results motivate OU dynamics as a core generative component and show that ecotype-informed structure can be incorporated without degrading predictive calibration, providing a principled route to ecological stratification.

3.7 Integration of ScPCA ecology with evolutionary modeling reveals a coupled eco-evolutionary structure in pediatric leukemias

Collectively, our results indicate that ScPCA-resolved tumor microenvironmental heterogeneity is associated with systematic differences in inferred evolutionary dynamics across pediatric leukemias. Rather than acting as unstructured background variation, microenvironmental context organizes patients into recurring ecological states that align with distinct regions of OU parameter space.

Three findings support this conclusion. First, ScPCA-derived TME structure defines immune ecotypes that correspond to characteristic evolutionary regimes, as reflected by clustering in (μ, $log_{10}\theta$) space. Second, inferred selection dynamics vary across ecological contexts: immune-enriched microenvironments tend to exhibit stronger stabilizing selection, whereas stromal/unknown-enriched environments show weaker mean reversion and more drift-dominated behavior. Third, diagnostic labels alone do not recapitulate the full structure of parameter variation, consistent with microenvironmental state providing complementary organization beyond diagnosis.

Together, integrating ScPCA-derived ecological states with ecotype-informed OU inference provides a unified eco-evolutionary framework for interpreting patient-specific evolutionary regimes in pediatric leukemia and offers an interpretable link between microenvironmental structure and stochastic evolutionary dynamics.

3.8 Flagship case study: ecotype-informed OU dynamics are interpretable, uncertainty-aware, and robust

To illustrate the interpretability and robustness of the ecotype-informed OU framework, we present a flagship case study integrating TME composition, ecotype assignments, and inferred OU parameters (Fig. 6). Ecotype membership was imbalanced in this cohort, with one dominant ecotype and several smaller groups (Fig. 6A, left), motivating hierarchical partial pooling while still enabling ecotype-stratified comparisons. Ecotype-level mean z-scored TME covariates clarified the ecological interpretation of each state, separating immune-enriched from stromal/other-enriched contexts (Fig. 6A, right).

Consistent with this ecological stratification, posterior summaries showed systematic differences in inferred OU parameters across ecotypes (Fig. 6B,C). Immune-enriched contexts tended to exhibit stronger stabilizing selection (higher $log_{10}\theta$) and more constrained trajectories, whereas stromal/other-enriched contexts corresponded to weaker stabilizing selection (lower $log_{10}\theta$), consistent with more drift-dominated dynamics. These patterns are presented as hypothesis-generating ecological signatures rather than outcome-linked conclusions.

Posterior predictive trajectories for representative patients indicated that the model captures time-ordered variation while maintaining calibrated uncertainty: posterior predictive medians tracked observed summaries, and 95% posterior predictive intervals reflected both measurement noise and latent stochasticity (Fig. 6D). Finally, varying the number of clusters (k=3–6) yielded qualitatively similar ecotype-associated trends in $log_{10}\theta$ (Fig. 7), supporting robustness of the ecological signal to ecotype granularity.

4. Discussion

In this study, we developed a tumor microenvironment (TME)–modulated Ornstein–Uhlenbeck (OU) modeling framework to integrate microenvironmental structure into evolutionary inference and applied it to pediatric leukemia using ScPCA single-cell cohorts. By combining single-cell–derived immune ecotypes with a stochastic model of stabilizing selection, we show that leukemic populations are associated with ecotype-specific evolutionary regimes rather than a single uniform dynamical process. These results highlight the TME as a structured ecological context that organizes inter-patient variation in inferred evolutionary dynamics in a reproducible and quantifiable manner.

A central finding is that immune ecotypes derived from TME composition map to distinct regions of OU parameter space, with systematic differences in inferred evolutionary optima (μ) and stabilizing selection strength (θ), while stochastic volatility is governed by a shared global diffusion scale σ_{proc}. Immune-enriched ecotypes tended to exhibit stronger stabilizing selection and more constrained dynamics, consistent with tighter mean reversion toward ecotype-associated optima. In contrast, stromal/unknown-enriched ecotypes showed weaker mean reversion and more drift-dominated behavior. Importantly, these patterns were not fully recapitulated by diagnosis labels alone, suggesting that microenvironmental state provides complementary organization beyond diagnostic classification.

From a modeling perspective, OU processes provide a natural continuous-time description of evolution under stabilizing selection, whereas classical branching or state-transition models emphasize diversification and clonal expansion. Our work is motivated by OU–Branching generative formulations that aim to unify stabilizing forces with lineage diversification; however, the analyses presented here focus on the ecotype-modulated OU component. Incorporating immune ecotypes as a hierarchical structure enables partial pooling across patients sharing similar ecological contexts, improving parameter stability while preserving patient-specific estimates and supporting interpretable comparisons across TME-defined regimes.

More broadly, this study contributes to a growing body of computational oncology research that emphasizes eco-evolutionary coupling as an organizing principle of tumor behavior. Rather than treating the TME as residual variation, our results support the view that discrete ecological states—derived in a data-driven manner from single-cell atlases—can be integrated directly into evolutionary models to reveal context-dependent constraints on tumor dynamics.

Limitations

Several limitations should be noted. First, the analysis relies primarily on cross-sectional single-cell datasets, with limited repeated sampling for most patients. Although the OU framework is formulated in continuous time, inference is constrained by the temporal resolution of available observations; thus, trajectories are inferred rather than directly observed over dense time courses. Second, immune ecotypes were derived from a reduced set of broad compositional TME features. While these capture major immune and stromal compartments, they do not encode finer-grained cellular states, spatial organization, or functional interactions such as cytokine signaling and cell–cell communication. Third, Bayesian inference provides uncertainty quantification and hierarchical pooling, but it introduces computational overhead that may limit scalability to very large cohorts or to richer hierarchical structures without approximation. Finally, pediatric leukemia has distinctive evolutionary and microenvironmental features; while the framework is general, the specific ecological regimes identified here may not transfer directly to other cancers without adaptation.

Future directions

The TME-modulated OU framework opens several avenues for future work. Incorporating longitudinal multi-sample datasets would enable explicit modeling of temporal changes in ecological state and potential regime switching, including microenvironmental remodeling during therapy, relapse, or immune escape. Methodologically, integrating explicit phylogenetic structure or clone-resolved lineage trees would enable the OU–Branching motivation to be fully instantiated, strengthening the link between stabilizing selection, branching diversification, and observed clonal architecture. Coupling the framework with genomic and epigenomic measurements may further clarify how genetic alterations interact with ecological constraints to shape evolutionary trajectories. Finally, the approach is not limited to leukemia: any setting where ecological context can be quantified—including solid tumors, microbial ecosystems, and developmental processes—may benefit from ecotype-modulated stochastic evolutionary models.

5. Conclusion

In summary, we present an ecotype-modulated OU modeling framework, motivated by OU–Branching formulations, that integrates tumor microenvironmental structure into evolutionary inference. Applied to pediatric leukemia, this approach identifies reproducible, ecotype-associated evolutionary regimes that are not fully explained by diagnosis alone. By coupling single-cell–resolved ecological context with stochastic evolutionary modeling and uncertainty-aware inference, our work provides a generalizable computational foundation for studying how microenvironmental structure shapes cancer evolution.

Supplementary Information

Supplementary Information accompanies this manuscript and includes Supplementary Methods, Supplementary Figures 1–4, Supplementary Tables 1–4, Supplementary Data 1–4, and Supplementary Notes 1–4.

Availability and Implementation

All analyses were implemented in Python 3.11. Source code and fully reproducible workflows (including environment specifications and figure-generation scripts) are available at https://github.com/shkim9391/TME_OU_Branching_Pediatric_Leukemia.

Data Availability Statement

The single-cell datasets analyzed in this study are publicly available from the Single-cell Pediatric Cancer Atlas (ScPCA) portal (projects SCPCP000008 and SCPCP000022) at https://scpca.alexslemonade.org/. Processed TME covariates, posterior samples, and analysis scripts are provided in the companion GitHub repository listed above. All derived data needed to reproduce the reported analyses are included in the repository and/or Supplementary Materials.

Ethics Statement

This study is a secondary analysis of de-identified, publicly available single-cell transcriptomic data from ScPCA and associated repositories. No new participants were recruited, and no identifiable private information was used; therefore, IRB/ethics committee approval and informed consent were not required for this analysis. Ethics review and participant consent were managed by the original studies and data repositories.

Funding

This research received no external funding.

Acknowledgements

We thank the ScPCA consortium and the contributing investigators for making these datasets publicly available. This acknowledgement does not imply endorsement of the findings.

Conflict of Interest

None declared.

Author contributions

S.-H.K.: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Data curation, Validation, Writing – review & editing

References

Aifantis I, Raetz E, Buonamici S. Molecular pathogenesis of T-cell leukaemia and lymphoma. Nat Rev Immunol. 2008;8(5):380–390.

Aktipis CA, Boddy AM, Gatenby RA et al. Life history trade-offs in cancer evolution. Nat Rev Cancer. 2013;13(12):883–892.

Andersson AK, Ma J, Wang J et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nat Genet. 2015;47:330–337.

Barkley D, Moncada R, Pour M et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. Nat Genet. 2022;54(8):1192–1201.

Baryawno N, Przybylski D, Kowalczyk MS et al. A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. Cell. 2019;177(7):1915–1932.e16.

Beaulieu JM, Jhwueng D-C, Boettiger C, O'Meara BC. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution. 2012;66(8):2369–2383.

Beerenwinkel N, Schwarz RF, Gerstung M, Markowetz F. Cancer evolution: mathematical models and computational inference. Syst Biol. 2015;64(1):e1–e25.

Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat Med. 2011;17(3):320–329.

Butler MA, King AA. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat. 2004;164(6):683–695.

Durrett R. Branching Process Models of Cancer. Springer, Cham; 2015. doi:10.1007/978-3-319-16065-8.

Gatenby RA, Gillies RJ. A microenvironmental model of carcinogenesis. Nat Rev Cancer. 2008;8(1):56–61.

Gatenby RA, Gillies RJ, Brown JS. Evolutionary dynamics of cancer prevention. Nat Rev Cancer. 2010;10(8):526–527.

Gillies RJ, Verduzco D, Gatenby RA. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. Nat Rev Cancer. 2012;12(7):487–493.

Granja JM, Klemm S, McGinnis LM et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat Biotechnol. 2019;37(12):1458–1465.

Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481:306–313.

Gröbner SN, Worst BC, Weischenfeldt J et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321–327.

Hansen TF. Stabilizing selection and the comparative analysis of adaptation. Evolution. 1997;51(5):1341–1351.

Hansen TF, Pienaar J, Orzack SH. A comparative method for studying adaptation to a randomly evolving environment. Evolution. 2008;62(8):1965–1977.

Huang S. Reprogramming cell fates: reconciling rarity with robustness. BioEssays. 2009;31(5):546–560.

Iacobucci I, Zeng AGX, Gao Q et al. Multipotent lineage potential in B cell acute lymphoblastic leukemia is associated with distinct cellular origins and clinical features. Nat Cancer. 2025;6(7):1242–1262. doi:10.1038/s43018-025-00987-2.

Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature. 2013;501:346–354.

Kather JN, Suarez-Carmona M, Charoentong P et al. Topography of cancer-associated immune cells in human solid tumors. eLife. 2018;7:e36967.

Kim S-H. Hybrid Ornstein–Uhlenbeck–Branching modeling of pediatric leukemia evolution: a computational extension and cohort-level application. bioRxiv. 2025. doi:10.1101/2025.11.07.686044.

Kim S-H. A hybrid Ornstein–Uhlenbeck–Branching framework unifies microbial and pediatric tumor evolution. Front Oncol. 2026;16:1727973. doi:10.3389/fonc.2026.1727973.

Kimmel JC, Hwang AB, Scaramozza A et al. Differentiation reveals latent features of aging and an energy barrier in murine myogenesis. Cell Rep. 2021;35(4):109046.

Ma X, Liu Y, Liu Y et al. Pan-cancer genome and transcriptome analyses of 1,699 pediatric leukemias and solid tumors. Nature. 2018;555(7696):371–376.

Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12(5):323–334.

McFarland CD, Korolev KS, Kryukov GV et al. Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci U S A. 2013;110(8):2910–2915.

Michor F, Hughes TP, Iwasa Y et al. Dynamics of chronic myeloid leukaemia. Nature. 2005;435(7046):1267–1270.

Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194(4260):23–28. doi:10.1126/science.959840.

Packer JS, Trapnell C. Single-cell multi-omics: an engine for new quantitative models of gene regulation. Trends Genet. 2018;34(9):653–665.

Puram SV, Tirosh I, Parikh AS et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell. 2017;171(7):1611–1624.e24.

Tirosh I, Izar B, Prakadan SM et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189–196.

Uyeda JC, Hansen TF, Arnold SJ, Pienaar J. The million-year wait for macroevolutionary bursts. Proc Natl Acad Sci U S A. 2011;108(38):15908–15913.

van Galen P, Hovestadt V, Wadsworth MH II et al. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. Cell. 2019;176(6):1265–1281.e24.

Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. Nat Biotechnol. 2016;34(11):1145–1160.

West J, Ma Y, Newton PK. Capitalizing on competition: an evolutionary model of competitive release in metastatic castration-resistant prostate cancer treatment. J Theor Biol. 2018;455:249–260.

Figure Captions

Figure 1. Overview of the hybrid OU–Branching modeling framework and its integration with tumor microenvironment (TME) ecotypes.
(A) Hybrid OU–Branching generative schematic: trait evolution along a lineage follows OU drift toward an ecotype-dependent optimum with branching events representing diversification. **(B)** Workflow: ScPCA single-cell data are summarized into TME composition features, used to assign participant-level immune ecotypes, and analyzed with an ecotype-informed OU model. Inference yields posteriors for patient-specific μ_i and θ_i, with stochastic volatility governed by a single shared global diffusion scale σ_{proc}. **(C)** Ecotype-informed priors: ecotype membership enters as a hierarchical structure (partial pooling) on μ_i and θ_i. **(D)** Posterior predictive check (PPC) for an exemplar patient-series, showing observed trait measurements versus the posterior predictive distribution (median and 95% predictive band). PPC coverage summarizes calibration; μ_i and θ_i vary by patient, while σ_{proc} is shared globally.

Figure 2. Global structure and variation of the tumor microenvironment (TME) across pediatric leukemia patients.
(A) Boxplots with overlaid patient-level jittered points showing the distribution of six TME-derived compositional features: T-cell, B-cell, myeloid, natural killer (NK) cell, stromal, and unknown/other fractions. Substantial inter-patient variability is observed across diagnostic categories. **(B)** Spearman correlation matrix of TME features highlighting strong positive associations among myeloid, stromal, and unknown compartments, as well as an inverse relationship between lymphoid (T/NK) and stromal programs. These patterns are consistent with coordinated immune-stromal coupling in leukemic bone marrow. **(C)** Principal component analysis (PCA) of TME

covariates illustrating major axes of microenvironmental variation. Diagnostic subgroups (e.g., DUX4-R, KMT2A-R, iAMP21, MEF2D-R, hyperdiploid) occupy distinct but overlapping regions of TME space, motivating diagnosis-informed ecological characterization. **(D)** K-means clustering elbow curve identifying k = 4 as the most parsimonious number of clusters, defining four reproducible TME immune ecotypes used for OU–Branching model calibration.

Figure 3. Immune ecotypes in pediatric leukemia.
(A) Principal component analysis (PCA) of tumor microenvironment (TME) composition across pediatric leukemia patients, colored by clinical diagnosis category: B-cell acute lymphoblastic leukemia, early T-cell precursor acute lymphoblastic leukemia, mixed phenotype acute leukemia, T-cell acute lymphoblastic leukemia, and unknown. The first two principal components (PC1 and PC2) capture the dominant axes of variation across T-cell, B-cell, myeloid, natural killer (NK), stromal, and unknown/other cell fractions. **(B)** Uniform manifold approximation and projection (UMAP) embedding of the same TME feature space, colored by data-driven immune ecotypes (Ecotypes 1-4) identified via K-means clustering. Ecotype 1 corresponds to T-cell-enriched profiles, Ecotype 2 to mixed immune compositions, Ecotype 3 to myeloid-skewed profiles, and Ecotype 4 to stromal/unknown-enriched microenvironments (see Panel D for detailed compositional signatures). **(C)** Mean TME composition for each immune ecotype, with colors indicating the fractional contributions of T cells, B cells, myeloid cells, NK cells, stromal cells, and unknown/other compartments. **(D)** Radar plots summarizing the characteristic immune profiles associated with each ecotype across the six TME components. Ecotype 1 is dominated by T-cell infiltration; Ecotype 2 exhibits moderate T-cell abundance with low myeloid and stromal content; Ecotype 3 displays strong myeloid polarization; and Ecotype 4 shows elevated stromal and unknown fractions. These immune ecotypes are used downstream as latent immunologic states in the OU–Branching evolutionary analyses.

Figure 4. OU dynamics stratified by immune ecotype.
(A) Posterior distributions of the OU drift mean (μ) stratified by immune ecotype. Violin plots display the full posterior density for each ecotype, with horizontal bars indicating posterior medians. **(B)** Same as in (A), with overlaid patient-level posterior means (jittered horizontally). Points are colored by clinical diagnosis (B-cell acute lymphoblastic leukemia [B-ALL], T-cell acute lymphoblastic leukemia [T-ALL], and acute myeloid leukemia [AML]), highlighting diagnosis-specific heterogeneity within ecotypes. **(C)** Posterior distributions of the stabilizing selection strength (θ), shown on a logarithmic (log_{10}) scale and stratified by immune ecotype. Violin plots depict the full posterior density, with the median indicated. **(D)** Same as in (C), with the patient-level posterior means overlaid and colored by clinical diagnosis. **(E)** Joint posterior structure of patient-level OU parameters, showing posterior mean drift (μ, x-axis) versus posterior mean stabilizing selection strength ($log_{10}\theta$, y-axis). Points are colored by immune ecotype and shaped by clinical diagnosis, with outlier patients labeled. Ecotype-level centroids (posterior mean ± one standard deviation in both dimensions) are shown as yellow markers with error bars, illustrating distinct OU dynamical regimes associated with each immune ecotype.

Figure 5. Ablation and baseline comparisons evaluate OU dynamics and ecotype-informed structure.
Predictive fit, point accuracy, and uncertainty calibration are shown for the full ecotype-informed OU model (TME-OU) and baseline/ablation alternatives: OU without ecotype/TME modulation (OU-only), shuffled ecotype labels (shuffled-TME), random-walk latent dynamics, and a static ecotype-only model. **(A)** Relative predictive fit measured by ΔELPD (PSIS-LOO) versus the OU-only baseline (error bars denote the standard error of ΔELPD from pointwise LOO contributions); inset shows a zoomed view of the near-zero ΔELPD region to highlight small differences among OU variants. **(B)** Point accuracy (RMSE) computed from posterior predictive means. **(C)** Empirical coverage of nominal 95% posterior predictive intervals (dashed line indicates the 0.95 target). Overall, OU mean-reverting dynamics substantially outperform the random-walk baseline, and the static ecotype-only model shows reduced point accuracy. Among OU variants, TME-OU and shuffled-TME are statistically indistinguishable from OU-only in ΔELPD within uncertainty, while maintaining near-nominal interval coverage.

Figure 6. Flagship case study: ecotype-informed OU dynamics are interpretable, uncertainty-aware, and robust.
(A) Cohort overview of immune ecotype composition and mean ScPCA-derived tumor microenvironment (TME) covariates (z-scored) defining each ecotype. **(B)** Schematic interpretation of ecotype-associated dynamical regimes: immune-rich contexts correspond to higher $log_{10}\theta$ (stronger stabilizing selection; more constrained trajectories near the drift target μ), whereas stromal/other-enriched contexts correspond to lower $log_{10}\theta$ (weaker stabilization; greater drift/dispersion). Hypothesis-generating; no outcome inference. **(C)** Posterior summaries of patient-level OU parameters stratified by ecotype, showing inferred drift/optimum μ and stabilizing selection strength $log_{10}\theta$ with uncertainty. **(D)** Representative longitudinal trajectories for exemplar patient-series (two participants, two series each), comparing observed ScPCA trait summaries with posterior predictive medians and 95% predictive bands.

Figure 7. Sensitivity of stabilizing selection estimates to ecotype resolution.
Posterior mean estimates of $log_{10}\theta$ (patient-level posterior mean, summarized within clusters) are shown for alternative *k*-means clusterings of the participant-level TME covariates with k = 3, 4, 5, and 6. Clusters (C1–Ck) are ordered by mean stromal z-score, and the diagonal annotations report cluster size and mean stromal z-score. Error bars indicate uncertainty (interval as plotted). Across resolutions, stromal-enriched clusters tend to exhibit lower $log_{10}\theta$ (weaker stabilizing selection), supporting robustness of the ecotype-associated trend.

Figure 1

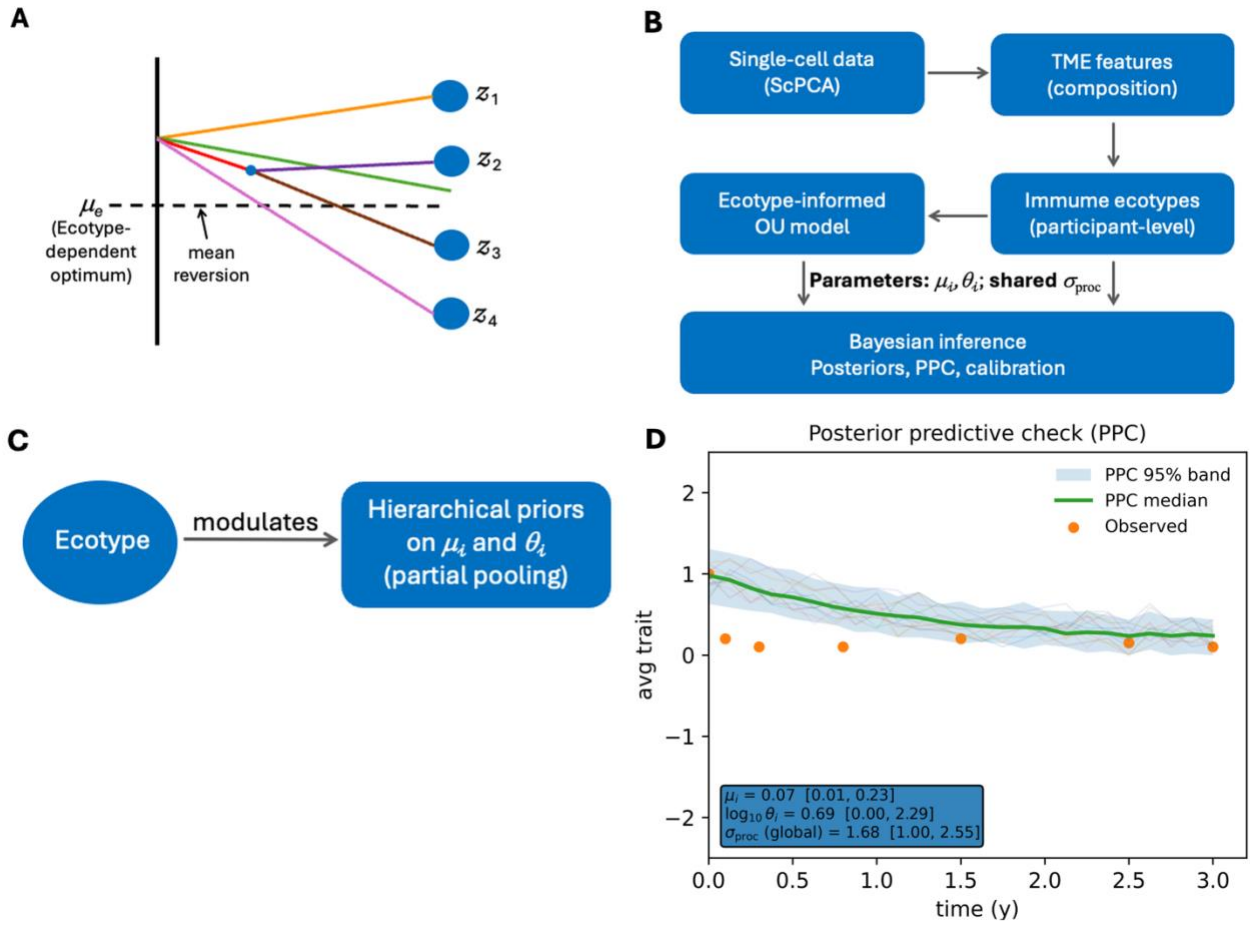

A
z_1
z_2
z_3
z_4
μ_e
(Ecotype-dependent optimum)
mean reversion
B
Single-cell data (ScPCA)
TME features (composition)
Immune ecotypes (participant-level)
Ecotype-informed OU model
Parameters: μ_i, θ_i; shared σ_{proc}
Bayesian inference
Posteriors, PPC, calibration
C
Ecotype
modulates
Hierarchical priors on μ_i and θ_i (partial pooling)
D
Posterior predictive check (PPC)
PPC 95% band
PPC median
Observed
avg trait
time (y)
μ_i = 0.07 [0.01, 0.23]
$\log_{10}\theta_i$ = 0.69 [0.00, 2.29]
σ_{proc} (global) = 1.68 [1.00, 2.55]

Figure 2

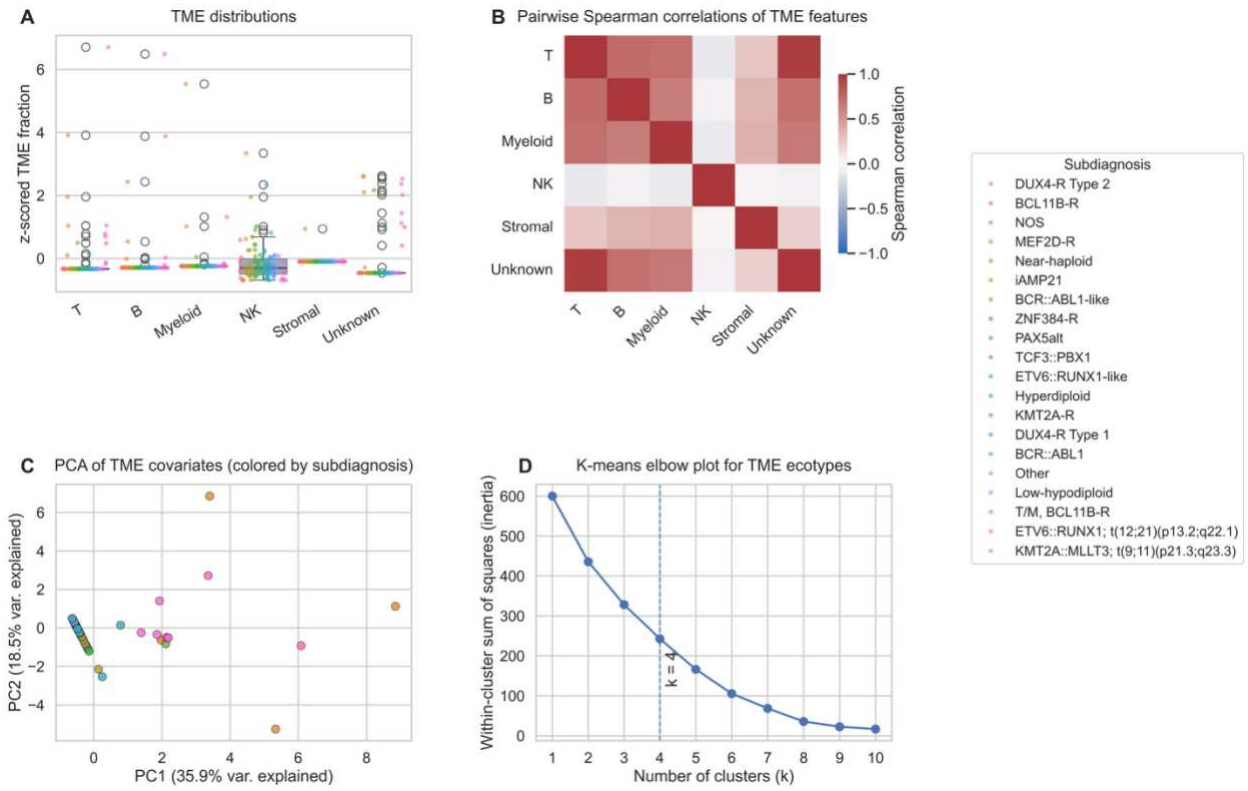

A
TME distributions
z-scored TME fraction
T
B
Myeloid
NK
Stromal
Unknown
B
Pairwise Spearman correlations of TME features
Spearman correlation
C
PCA of TME covariates (colored by subdiagnosis)
PC2 (18.5% var. explained)
PC1 (35.9% var. explained)
D
K-means elbow plot for TME ecotypes
Within-cluster sum of squares (inertia)
k = 4
Number of clusters (k)
Subdiagnosis
DUX4-R Type 2
BCL11B-R
NOS
MEF2D-R
Near-haploid
iAMP21
BCR::ABL1-like
ZNF384-R
PAX5alt
TCF3::PBX1
ETV6::RUNX1-like
Hyperdiploid
KMT2A-R
DUX4-R Type 1
BCR::ABL1
Other
Low-hypodiploid
T/M, BCL11B-R
ETV6::RUNX1; t(12;21)(p13.2;q22.1)
KMT2A::MLLT3; t(9;11)(p21.3;q23.3)

Figure 3

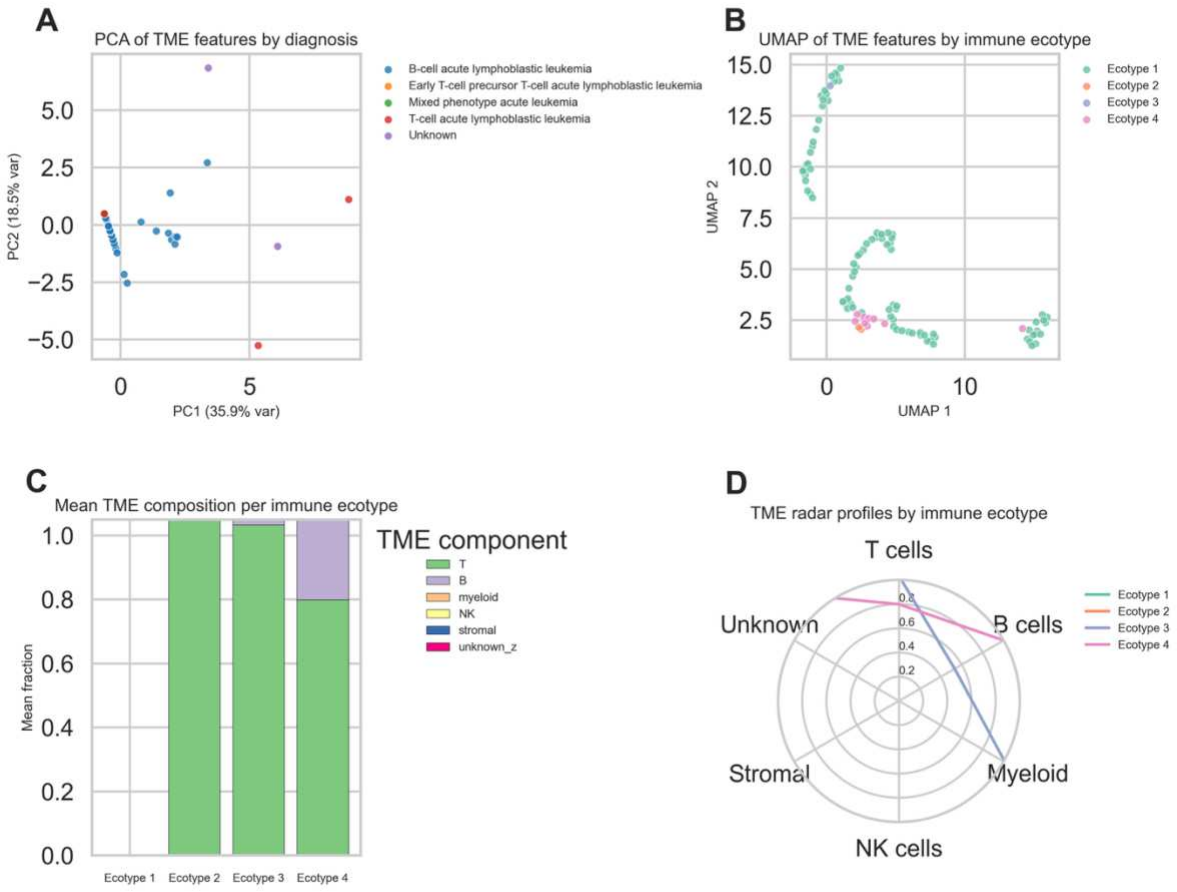

A
PCA of TME features by diagnosis
B-cell acute lymphoblastic leukemia
Early T-cell precursor T-cell acute lymphoblastic leukemia
Mixed phenotype acute leukemia
T-cell acute lymphoblastic leukemia
Unknown
PC2 (18.5% var)
PC1 (35.9% var)
B
UMAP of TME features by immune ecotype
Ecotype 1
Ecotype 2
Ecotype 3
Ecotype 4
UMAP 2
UMAP 1
C
Mean TME composition per immune ecotype
TME component
T
B
myeloid
NK
stromal
unknown_z
Mean fraction
Ecotype 1
Ecotype 2
Ecotype 3
Ecotype 4
D
TME radar profiles by immune ecotype
T cells
B cells
Myeloid
NK cells
Stromal
Unknown

Figure 4

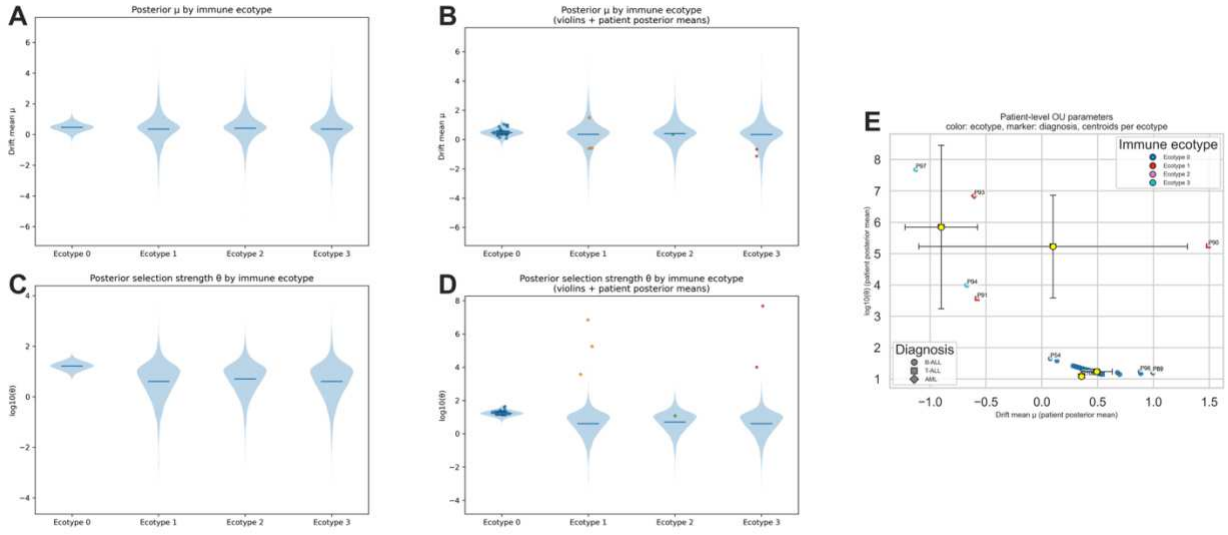

A
Posterior μ by immune ecotype
Drift mean μ
Ecotype 0
Ecotype 1
Ecotype 2
Ecotype 3
B
Posterior μ by immune ecotype
(violins + patient posterior means)
Drift mean μ
Ecotype 0
Ecotype 1
Ecotype 2
Ecotype 3
C
Posterior selection strength θ by immune ecotype
log10(θ)
Ecotype 0
Ecotype 1
Ecotype 2
Ecotype 3
D
Posterior selection strength θ by immune ecotype
(violins + patient posterior means)
log10(θ)
Ecotype 0
Ecotype 1
Ecotype 2
Ecotype 3
E
Patient-level OU parameters
color: ecotype, marker: diagnosis, centroids per ecotype
Immune ecotype
Ecotype 0
Ecotype 1
Ecotype 2
Ecotype 3
Diagnosis
B-ALL
T-ALL
AML
Drift mean μ (patient posterior mean)

Figure 5

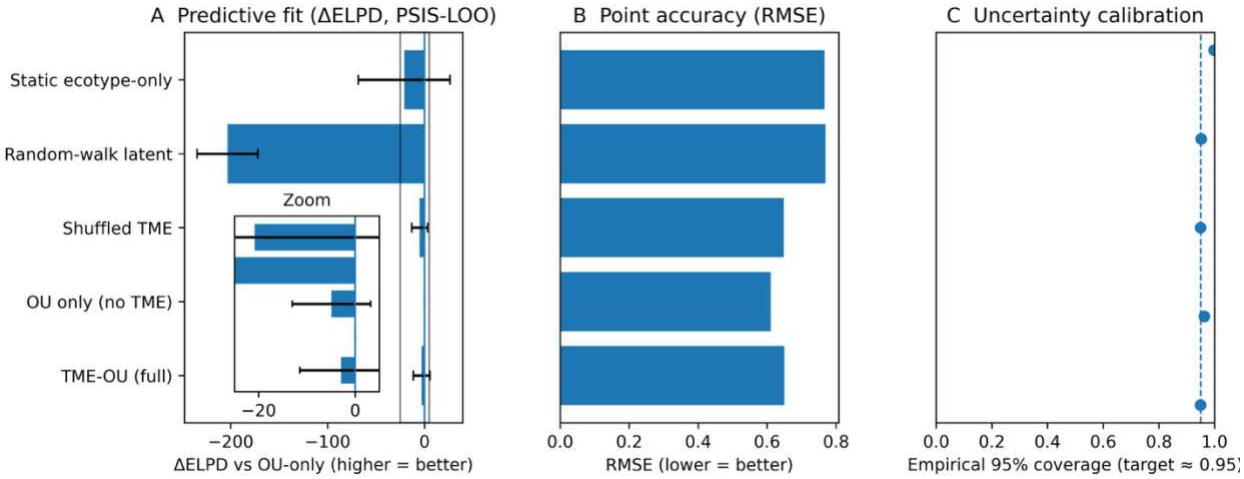

A Predictive fit (ΔELPD, PSIS-LOO)
B Point accuracy (RMSE)
C Uncertainty calibration
Static ecotype-only
Random-walk latent
Zoom
Shuffled TME
OU only (no TME)
TME-OU (full)
−20
0
−200
−100
0
ΔELPD vs OU-only (higher = better)
0.0
0.2
0.4
0.6
0.8
RMSE (lower = better)
0.0
0.2
0.4
0.6
0.8
1.0
Empirical 95% coverage (target ≈ 0.95)

Figure 6

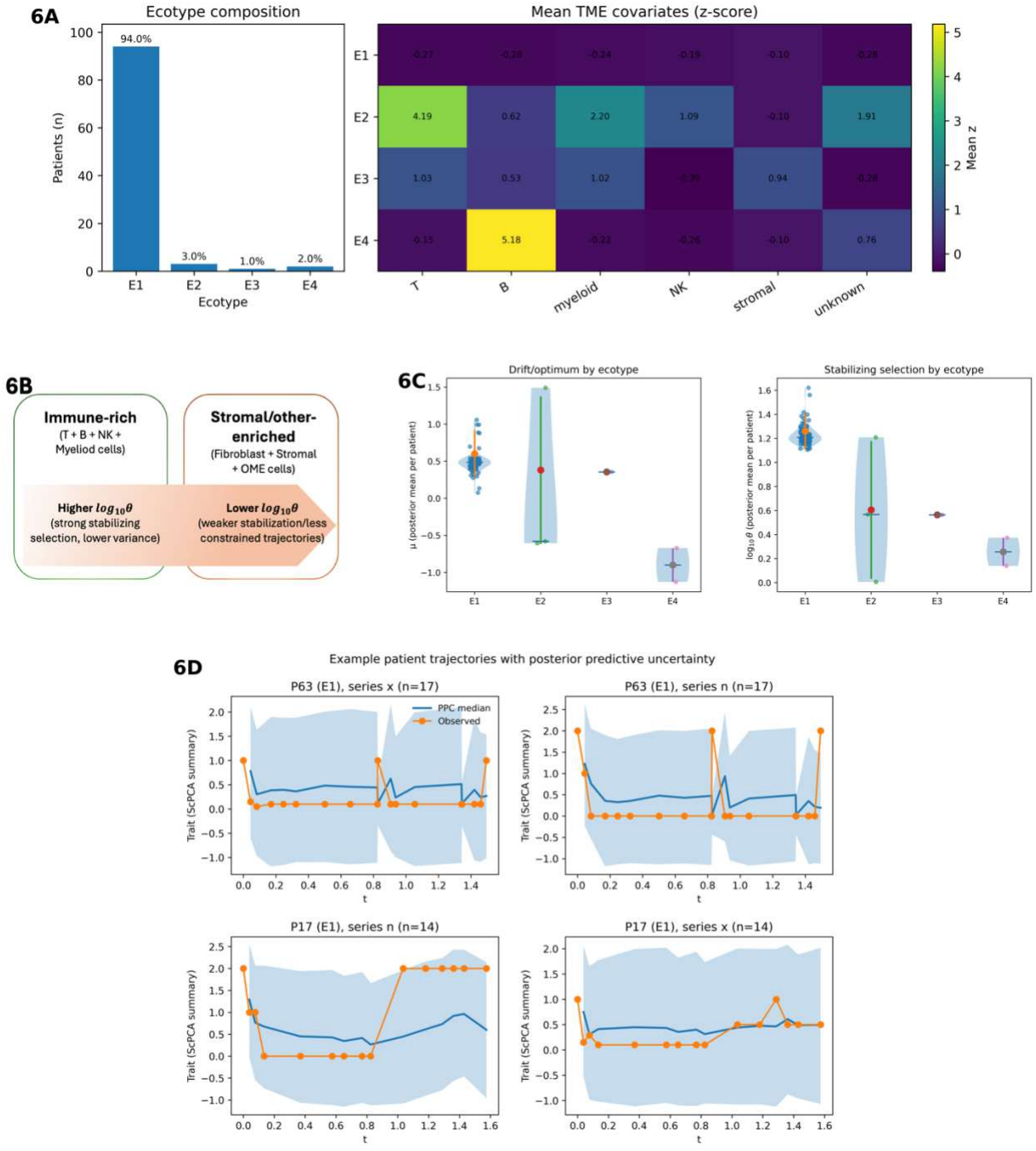
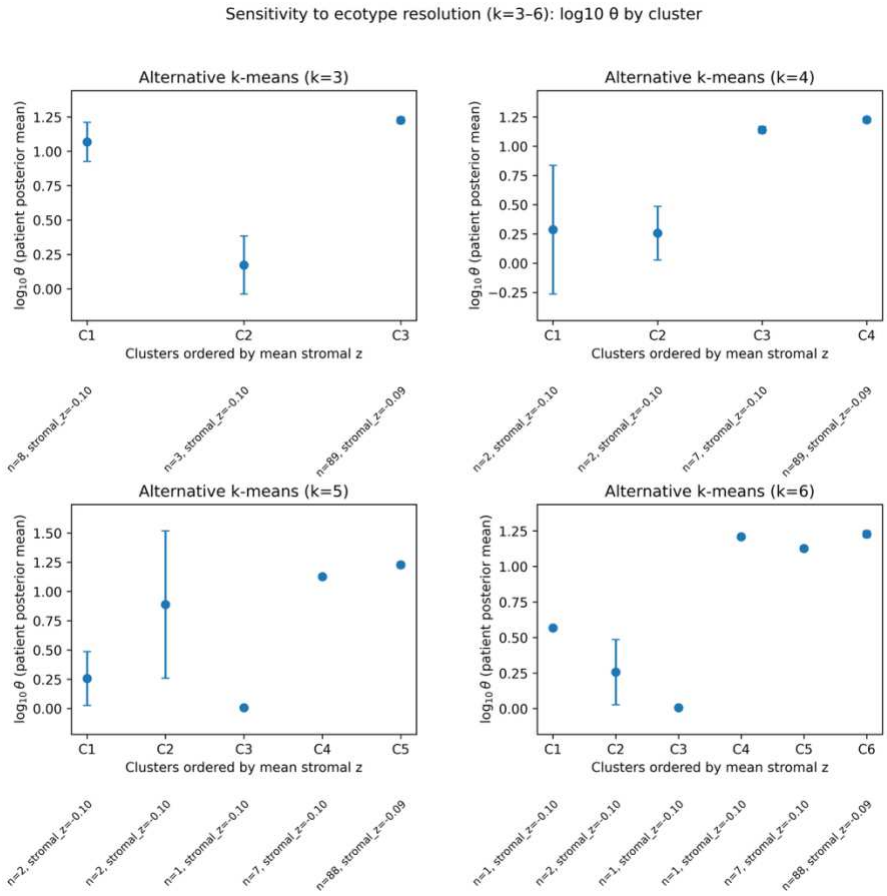

6A
Ecotype composition
Patients (n)
Ecotype
E1
E2
E3
E4
94.0%
3.0%
1.0%
2.0%
Mean TME covariates (z-score)
T
B
myeloid
NK
stromal
unknown
Mean z
6B
Immune-rich
(T + B + NK + Myeliod cells)
Stromal/other-enriched
(Fibroblast + Stromal + OME cells)
Higher $log_{10}\theta$
(strong stabilizing selection, lower variance)
Lower $log_{10}\theta$
(weaker stabilization/less constrained trajectories)
6C
Drift/optimum by ecotype
μ (posterior mean per patient)
Stabilizing selection by ecotype
$log_{10}\theta$ (posterior mean per patient)
6D
Example patient trajectories with posterior predictive uncertainty
P63 (E1), series x (n=17)
P63 (E1), series n (n=17)
P17 (E1), series n (n=14)
P17 (E1), series x (n=14)
PPC median
Observed
Trait (ScPCA summary)
t

Figure 7


Sensitivity to ecotype resolution (k=3–6): log10 θ by cluster
Alternative k-means (k=3)
Alternative k-means (k=4)
Alternative k-means (k=5)
Alternative k-means (k=6)
log10 θ (patient posterior mean)
Clusters ordered by mean stromal z
C1
C2
C3
C4
C5
C6
n=8, stromal_z=-0.10
n=3, stromal_z=-0.10
n=89, stromal_z=-0.09
n=2, stromal_z=-0.10
n=2, stromal_z=-0.10
n=7, stromal_z=-0.10
n=89, stromal_z=-0.09
n=2, stromal_z=-0.10
n=2, stromal_z=-0.10
n=1, stromal_z=-0.10
n=7, stromal_z=-0.10
n=88, stromal_z=-0.09
n=1, stromal_z=-0.10
n=2, stromal_z=-0.10
n=1, stromal_z=-0.10
n=1, stromal_z=-0.10
n=7, stromal_z=-0.10
n=88, stromal_z=-0.09

Supplementary Files

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

- SupplementaryInformation.pdf
- patientmastertable.csv
- covariatematrix.csv
- SupplementaryData3ecotypeposteriorsummaries.csv
- SupplementaryData4patientposteriorsummaries.csv