Additional file 1
Supplementary Information

Continuous immune composition retains longitudinal information beyond coarse ecological-context labels in pediatric leukemia

Seung-Hwan Kim[1,2*]

[1]Department of Biology, Fisher College, Boston, MA, USA

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

[*]Correspondence: seung-hwan.kim@fisher.edu

This file contains Supplementary Methods, Supplementary Figures S1–S6, Supplementary Tables S1–S9, descriptions of Supplementary Data 1–8, and Supplementary Notes 1–6 supporting single-cell annotation harmonization, candidate ecological-context discovery and robustness assessment, Bayesian Ornstein–Uhlenbeck (OU) model implementation, predictive validation and sensitivity analyses, parameter recovery, posterior diagnostics, posterior predictive calibration, and interpretation of tumor microenvironment (TME) ecological composition.

Supplementary Methods

ScPCA data processing and common cell-type harmonization

Single-cell pediatric leukemia data and associated metadata were obtained from the Single-cell Pediatric Cancer Atlas (ScPCA) Portal. Two public leukemia projects, SCPCP000008 and SCPCP000022, were included. Because the projects differed in available cell-type annotations and annotation vocabularies, sample-level cellular composition was regenerated using a common cell-type mapping workflow rather than directly combining previously derived project-specific feature tables.

For each sample, available single-cell annotations were mapped to harmonized broad cellular compartments using the same mapping rules across projects. The primary cross-project ecological analysis was restricted to three consistently represented immune compartments: combined T-cell/natural-killer (TNK), B-lineage, and myeloid cells. Unknown/unassigned, excluded, other, blast/malignant, progenitor, erythroid, and stromal-associated categories were not included in the primary three-component composition. Stromal-associated cells were excluded from the primary simplex because this compartment was essentially absent in SCPCP000008 and near zero in SCPCP000022, preventing a comparable cross-project representation.

For sample s, the retained compartment fractions were renormalized as

$$p_{s,j} = \frac{n_{s,j}}{n_{s,\mathrm{TNK}} + n_{s,\mathrm{B}} + n_{s,\mathrm{Myeloid}}},$$

where $n_{s,j}$ is the number of cells mapped to compartment j. Thus,

$$p_{s,\mathrm{TNK}} + p_{s,\mathrm{B}} + p_{s,\mathrm{Myeloid}} = 1.$$

Sample-level compositions were aggregated to the participant level so that participants represented by multiple samples did not receive disproportionate weight during ecological-context discovery. Participants with zero total abundance across the three retained compartments were excluded from the primary simplex. The common-mapper workflow regenerated profiles for 113 participants, of whom 104 satisfied eligibility criteria for the primary TNK/B/myeloid composition (Supplementary Fig. S1A).

Participant reconciliation and longitudinal support

ScPCA participant identifiers were reconciled with the legacy longitudinal P1–P100 identifiers using a curated participant crosswalk. Of the 113 regenerated participants, 100 could be mapped to the previous P1–P100 cohort. Ninety-four participants were both eligible for the

primary TNK/B/myeloid simplex and mapped to the previous cohort. Within B-cell acute lymphoblastic leukemia (B-ALL), 99 participants satisfied primary compositional eligibility and 90 were additionally mapped to the previous cohort (Supplementary Fig. S1A).
Longitudinal modeling was restricted further according to endpoint-specific observation support. For the primary x endpoint, 34 participants contributed at least one valid positive-time transition. These participants contributed 241 valid observations and 207 transitions. The median number of unique observation times was 6 per participant (range, 2–18), and median follow-up was 1.283 years. Follow-up duration and observation density varied substantially across participants (Supplementary Fig. S1B), motivating explicit use of the irregular-time OU transition density.

Raw compositional representation and centered log-ratio sensitivity analysis

The primary ecological representation used the raw renormalized TNK/B/myeloid proportions. Because these variables form a constrained three-part composition, candidate ecological-context discovery was additionally repeated using a centered log-ratio (CLR) representation.
A pseudocount of 10^{-6} was added to each component to permit logarithmic transformation of zero-valued fractions. The composition was then renormalized:

$$p^*_{ij} = \frac{p_{ij} + 10^{-6}}{\sum_{k=1}^{3}(p_{ik} + 10^{-6})},$$

and the CLR coordinates were calculated as

$$\mathrm{clr}(p^*_{ij}) = \log p^*_{ij} - \frac{1}{3}\sum_{k=1}^{3} \log p^*_{ik}.$$

The resulting CLR coordinates were standardized to zero mean and unit variance across participants before Euclidean K-means clustering. In contrast, the raw-composition analysis used the renormalized TNK/B/myeloid proportions directly without additional standardization.
Cluster labels C1–Ck were assigned deterministically after fitting to facilitate comparisons across analyses. Labels were ordered according to mean composition on the original biological scale, first by decreasing TNK fraction, then by decreasing B fraction, and then by decreasing myeloid fraction. This ordering affected only cluster naming and did not affect K-means fitting.

Candidate ecological-context discovery and resolution scan

Candidate ecological contexts were treated as descriptive coarse-grainings of continuous participant-level immune composition rather than as predefined leukemia subtypes. K-means clustering was evaluated at

$$k = 2, 3, 4, 5, 6$$

for both the raw-composition and standardized CLR representations. K-means fitting used 100 random initializations with a fixed random seed.

For each resolution, we recorded the minimum and maximum cluster sizes, coefficient of variation of cluster sizes, silhouette coefficient, and association of cluster membership with source project and diagnosis. Associations with categorical variables were quantified using Cramér's V.

The raw-composition $k = 2$ solution was retained as the conservative working coarse-graining because it combined moderate cluster separation with comparatively balanced group sizes and limited association with project or diagnosis. The two groups were subsequently described according to their composition as TNK-dominant and B/Myeloid-shifted candidate ecological contexts.

Cluster-resolution sensitivity is summarized in Supplementary Fig. S2. Increasing k in the raw representation generated progressively smaller groups, while the CLR representation produced high silhouette coefficients at several resolutions but strongly imbalanced solutions, including groups containing only two participants at higher k. These results motivated interpretation of the $k = 2$ partition as a descriptive low-resolution summary rather than evidence for uniquely defined discrete ecological states.

Bootstrap assessment of ecological-context stability

Cluster stability was evaluated using participant-level bootstrap resampling. For each bootstrap replicate, n participants were sampled with replacement from the n eligible participants and K-means was refitted to the resampled composition matrix. The resulting bootstrap centroids were then used to predict cluster membership for all participants in the original dataset.

Stability was evaluated relative to the reference clustering solution using pairwise co-clustering agreement and cluster-specific Jaccard similarity. Global pairwise agreement was calculated by comparing whether each pair of participants was assigned to the same or different clusters in the reference and bootstrap solutions. For each reference cluster, the bootstrap-predicted cluster with maximum overlap was identified and Jaccard similarity was calculated as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Five hundred bootstrap replicates were used. Stability distributions were summarized by their mean, median, and empirical 5th and 95th percentiles.

The same analysis was repeated after restricting the cohort to B-ALL to reduce confounding by broad leukemia lineage. Among 99 eligible B-ALL participants, the raw $k = 2$ solution contained 73 TNK-dominant and 26 B/Myeloid-shifted participants and retained strong resampling stability with minimal association with source project.

Concordance between raw and CLR ecological-context assignments

Representation sensitivity was evaluated by comparing raw-composition and CLR-derived K-means assignments at each candidate resolution. Concordance was quantified using the adjusted Rand index (ARI),

$$\mathrm{ARI} = \frac{\mathrm{RI} - \mathbb{E}[\mathrm{RI}]}{1 - \mathbb{E}[\mathrm{RI}]},$$

where RI denotes the Rand index and the expectation is taken under random assignment with the observed cluster sizes.

Raw-versus-CLR concordance was low across $k = 2$–6 despite strong within-representation separation for several CLR solutions. Consequently, the revised analysis distinguishes reproducibility within a chosen representation from invariance of discrete ecological labels across representations. Continuous immune composition was therefore retained as the primary ecological information for longitudinal modeling, while the $k = 2$ labels were evaluated separately as a coarse categorical representation.

Construction of the primary longitudinal x endpoint

The primary longitudinal outcome was the prespecified semiquantitative leukemia-state variable x contained in the longitudinal `Series` table. The x and n series were separated before modeling and were not pooled into a common Gaussian likelihood.

Within each participant, x observations were ordered by time. Duplicate participant-time rows were audited explicitly. Exact duplicates were collapsed, and same-time groups containing multiple coded values were resolved according to the predefined preprocessing rules before transitions were constructed.

For consecutive valid observations,

$$(x_{i,j-1}, t_{i,j-1}) \quad \text{and} \quad (x_{i,j}, t_{i,j}),$$

the transition variables were defined as

$$x_{\mathrm{prev},ij} = x_{i,j-1},$$

$$x_{ij} = x_{i,j},$$

and

$$\Delta t_{ij} = t_{i,j} - t_{i,j-1}.$$

Only transitions satisfying $\Delta t_{ij} > 0$ were retained. The final primary model-input table contained 207 transitions from 34 participants.

Continuous ecological covariates and coarse context variable

Each longitudinally supported participant was linked to the participant-level immune composition derived independently from the ScPCA common-mapper analysis. Continuous ecological covariates consisted of B and myeloid fractions. Because TNK, B, and myeloid fractions sum to one, all three components were not entered simultaneously as unconstrained regression predictors; TNK therefore serves as the implicit compositional reference when B and myeloid are included jointly.

For full-data modeling, ecological predictors were standardized across participants as

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}.$$

The coarse $k = 2$ ecological context was represented as a binary participant-level predictor distinguishing TNK-dominant from B/Myeloid-shifted composition. Among the 34 longitudinally supported participants, 27 were TNK-dominant and 7 were B/Myeloid-shifted, contributing 169 and 38 transitions, respectively.

Revised Ornstein–Uhlenbeck transition model

The primary longitudinal model treated the scalar x endpoint as an effective Ornstein–Uhlenbeck (OU) stochastic process,

$$dX_i(t) = \theta_i\big(\mu_i - X_i(t)\big)\,dt + \sigma_{\mathrm{proc}}\,dW_i(t),$$

where μ_i is the participant-specific effective attractor, $\theta_i > 0$ is the effective mean-reversion strength, $\sigma_{\mathrm{proc}} > 0$ is the shared process-diffusion scale, and $W_i(t)$ is a standard Wiener process.

Rather than introducing a separate Gaussian measurement-error layer, the revised analysis used the exact OU transition density directly for consecutive observed x values:

$$X_i(t+\Delta t) \mid X_i(t) \sim \mathcal{N}\left[\mu_i + \big(X_i(t)-\mu_i\big)e^{-\theta_i \Delta t},\ \frac{\sigma^2_{\text{proc}}}{2\theta_i}\left(1-e^{-2\theta_i \Delta t}\right)\right].$$

This likelihood accommodates irregular observation intervals directly. The OU formulation is used as an effective stochastic description of the coded longitudinal endpoint; mean reversion is not interpreted as direct evidence of stabilizing evolutionary selection.

Ecological modulation of OU parameters

Ecological predictors were allowed to modify both the effective attractor and mean-reversion strength. For participant i,

$$\mu_i = \mu_0 + \mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta}_\mu,$$

and

$$\log \theta_i = \theta_0 + \mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta}_\theta,$$

such that

$$\theta_i = \exp\left(\theta_0 + \mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta}_\theta\right).$$

Here, $\mathbf{X}_i$ denotes the standardized ecological predictor vector for participant i. Depending on the ablation model, this vector contained B composition, myeloid composition, the coarse ecological-context indicator, or combinations thereof.
These coefficients quantify statistical associations between participant-level ecological composition and the parameters of the effective OU model. They are not interpreted as causal effects of specific immune compartments on leukemia dynamics.

Matched OU ablation models

Seven matched models were fitted to distinguish continuous ecological information from discrete context effects and model flexibility:

$$M_0: \quad \text{OU only},$$

$$M_B: \quad \text{OU + B},$$

$$M_M : \quad \text{OU + Myeloid},$$

$$M_E : \quad \text{OU + B + Myeloid},$$

$$M_C : \quad \text{OU + context},$$

$$M_{EC} : \quad \text{OU + B + Myeloid + context},$$

and

$$M_S : \quad \text{OU + shuffled context}.$$

For M_S, context labels were permuted between participants while preserving the observed number of participants in each context. Shuffling was performed at the participant level so that all transitions from a participant retained the same shuffled label.
All models used the same OU transition likelihood and corresponding prior families. Thus, model comparisons isolate differences in ecological predictor structure rather than differences in the underlying stochastic likelihood.

Bayesian priors, posterior sampling, and convergence assessment

The reference prior specification was

$$\mu_0 \sim \mathcal{N}(0, 1),$$

$$\theta_0 \sim \mathcal{N}(0, 1),$$

$$\sigma_{\text{proc}} \sim \text{HalfNormal}(1),$$

and

$$\beta_{\mu,j}, \beta_{\theta,j} \sim \mathcal{N}(0, 0.35).$$

Posterior inference was performed using Hamiltonian Monte Carlo with the No-U-Turn Sampler (NUTS) implemented in `PyMC`. Final full-data ablation models used four chains, 3,000 tuning iterations, and 2,000 retained posterior draws per chain with `target_accept=0.99`.

Posterior-predictive draws and pointwise log-likelihood values were retained for every fitted model.

Convergence was evaluated using divergent-transition counts, rank-normalized $\hat{R}$, and bulk and tail effective sample sizes. All seven primary full-data ablation models completed with zero divergent transitions, maximum $\hat{R}$ approximately 1.00, and high effective sample sizes.

Transition-level PSIS-LOO diagnostics and influential observations

Pointwise log-likelihood values were used to calculate Pareto-smoothed importance-sampling leave-one-out cross-validation (PSIS-LOO) as a transition-level diagnostic. Expected log predictive density (ELPD), effective number of parameters, and Pareto-k values were recorded for each model.

Because multiple transitions were contributed by the same participant, transition-level PSIS-LOO was not treated as the primary measure of generalization. In addition, several models showed elevated Pareto-k values driven principally by the same abrupt P63 transition from $x = 0.10$ to $x = 1.00$ over $\Delta t \approx 0.003$ years (Supplementary Fig. S6A). A P56 transition from $x = 0.05$ to $x = 1.00$ over $\Delta t = 0.846$ years was additionally influential in the ecology-plus-context model.

These observations were retained rather than excluded. Their influence was used to motivate explicit participant-level model refitting rather than reliance on transition-level importance-sampling approximations.

True leave-one-participant-out validation

Predictive generalization was evaluated using true leave-one-participant-out (LOPO) validation. For each of the 34 longitudinally supported participants, all transitions from that participant were removed from the training data and the model was refitted to the remaining participants.

Within each fold, ecological predictor means and standard deviations were calculated using training participants only. The resulting scaling parameters were then applied to the held-out participant, preventing information leakage through predictor standardization.

For held-out transition j, posterior-predictive log density was calculated by integrating the transition likelihood over posterior draws,

$$\log p(y_j \mid y_{-i}) = \log \left[\frac{1}{S} \sum_{s=1}^{S} p(y_j \mid \Theta^{(s)}, y_{-i}) \right],$$

where y_{-i} denotes the training data excluding participant i, $\Theta^{(s)}$ is posterior draw s, and S is the number of retained posterior samples.

Transition-level log predictive densities were summed within each held-out participant to obtain a participant-level ELPD contribution. Aggregate model performance was summarized by the total held-out ELPD across participants, the mean participant-level ELPD, its standard error across participants, and the mean log predictive density per held-out transition. Paired model comparisons were performed at the participant level. For models A and B, the participant-specific predictive difference was

$$\Delta_i = \mathrm{ELPD}_{i,A} - \mathrm{ELPD}_{i,B}.$$

The mean paired difference, total paired difference, standard error of the participant-level differences, and numbers of participants favoring each model were recorded. This paired analysis was used to distinguish a modest aggregate predictive advantage from uniform participant-level superiority.

LOPO refits used two chains, 1,500 tuning iterations, and 1,000 retained posterior draws per chain with `target_accept=0.99`. Fold-specific sampler diagnostics, including divergences, maximum $\hat{R}$, and minimum bulk ESS, were retained. No LOPO fold showed divergent transitions in the principal comparisons.

Transformed-x sensitivity analysis

Because the primary x endpoint is bounded and semiquantitative, sensitivity to endpoint scale and boundary treatment was evaluated using a clipped logit transformation. For clipping parameter ϵ,

$$x^* = \mathrm{clip}\left(x, \epsilon, 1-\epsilon\right),$$

followed by

$$x_{\mathrm{logit}} = \log\left(\frac{x^*}{1-x^*}\right).$$

The same transformation was applied to both the preceding and subsequent state in each transition so that the OU likelihood remained internally consistent on a single transformed state space.

Three clipping values were evaluated:

$$\epsilon = 0.01, \quad 0.025, \quad 0.05.$$

The transformed datasets retained the same 34 participants and 207 transitions as the original-scale analysis. Across these settings, 33 current-state values and 50 preceding-state

values were clipped at the upper boundary; no values required lower-bound clipping in the final transition tables.

For each ϵ, four full-data OU models were refitted:

$$X_0: \quad \text{OU only},$$

$$X_B: \quad \text{OU + B},$$

$$X_E: \quad \text{OU + B + Myeloid},$$

and

$$X_C: \quad \text{OU + context}.$$

These models used the same OU transition formulation and prior families as the corresponding original-scale models. Posterior-predictive draws and pointwise log-likelihood values were retained for every fit.

The B-associated effect on the effective attractor remained negative across all three clipping values (Supplementary Fig. S6B). Because $\epsilon = 0.05$ maps the effective upper and lower bounds to a symmetric finite logit range and produced the same qualitative inference as the smaller clipping values, this setting was selected for true participant-level LOPO sensitivity analysis.

True LOPO at $\epsilon = 0.05$ followed the same procedure as the original-scale analysis, including training-only predictor standardization and complete refitting for each held-out participant. Models X_0, X_B, X_E, and X_C were compared using paired participant-level ELPD differences.

Secondary ordinal n endpoint

The longitudinal n series was analyzed separately from x. Because n takes the ordered values 0, 1, and 2, it was treated as a three-level ordinal leukemia-state indicator rather than as a continuous variable or literal clone count.

For consecutive observations, transitions were defined as

$$(n_{\text{prev}}, n, \Delta t).$$

The same 34 participants contributed 207 valid n transitions. The next-state distribution contained 115 observations in state 0, 12 in state 1, and 80 in state 2. The observed transition matrix was

	$n=0$	$n=1$	$n=2$
$n_{\text{prev}}=0$	79	0	17
$n_{\text{prev}}=1$	9	1	2
$n_{\text{prev}}=2$	27	11	61

.

The next ordinal state was modeled using an ordered-logistic likelihood,

$$n_j \sim \text{OrderedLogistic}\left(\eta_j, \mathbf{c}\right),$$

where η_j is the linear predictor and

$$\mathbf{c} = (c_1, c_2), \qquad c_1 < c_2,$$

are the cutpoints separating the three outcome categories.
The cutpoints were parameterized as

$$c_1 \sim \mathcal{N}(-0.75, 1),$$

$$\delta_c \sim \text{HalfNormal}(1),$$

and

$$c_2 = c_1 + \delta_c,$$

which enforces the required ordering without use of a non-default transformed initial-value specification.
The baseline linear predictor included the standardized previous state and elapsed interval,

$$\eta_j = \beta_{\text{prev}} n^{(z)}_{\text{prev},j} + \beta_{\Delta t} \Delta t^{(z)}_j.$$

Matched ecological ablations were then defined as

$$N_0: \quad n_{\text{prev}} + \Delta t,$$

$$N_B: \quad N_0 + B,$$

$$N_E: \quad N_0 + B + \text{Myeloid},$$

$$N_C : \quad N_0 + \text{Context},$$

and

$$N_{EC} : \quad N_0 + B + \text{Myeloid} + \text{Context}.$$

Regression coefficients were assigned regularizing priors,

$$\beta_j \sim \mathcal{N}(0, 0.5).$$

Models were fitted using four chains, 3,000 tuning iterations, and 2,000 retained draws per chain with `target_accept=0.99`. Posterior-predictive samples and pointwise log-likelihood values were retained for all five models. All ordinal models completed with zero divergent transitions, maximum $\hat{R}$ approximately 1.00, and high bulk and tail effective sample sizes. Transition-level PSIS-LOO was used as a secondary comparison diagnostic. Because state 1 was uncommon and the ordinal analysis was intended as a consistency analysis rather than an independent prediction model, no participant-level LOPO refitting was performed for this secondary endpoint.

Prior-sensitivity analysis

Sensitivity of the principal OU conclusions to prior specification was evaluated for the four core original-scale models M_0, M_B, M_E, and M_C. The observed transition data, OU likelihood, ecological predictor definitions, and model architecture were held fixed while prior scales were varied.

Three prior regimes were considered: narrow, reference, and wide. Under the narrow regime,

$$\mu_0 \sim \mathcal{N}(0, 0.5),$$

$$\theta_0 \sim \mathcal{N}(0, 0.5),$$

$$\sigma_{\text{proc}} \sim \text{HalfNormal}(0.5),$$

and

$$\beta_{\mu,j}, \beta_{\theta,j} \sim \mathcal{N}(0, 0.20).$$

The reference regime used

$$\mu_0 \sim \mathcal{N}(0, 1),$$

$$\theta_0 \sim \mathcal{N}(0, 1),$$

$$\sigma_{\text{proc}} \sim \text{HalfNormal}(1),$$

and

$$\beta_{\mu,j}, \beta_{\theta,j} \sim \mathcal{N}(0, 0.35).$$

Under the wide regime,

$$\mu_0 \sim \mathcal{N}(0, 2),$$

$$\theta_0 \sim \mathcal{N}(0, 1.5),$$

$$\sigma_{\text{proc}} \sim \text{HalfNormal}(2),$$

and

$$\beta_{\mu,j}, \beta_{\theta,j} \sim \mathcal{N}(0, 0.70).$$

Each prior-sensitivity model was fitted using four chains, 2,000 tuning iterations, and 1,500 retained posterior draws per chain with `target_accept=0.99`. Posterior means, 95% HDIs, posterior directional probabilities, divergent-transition counts, $\hat{R}$, effective sample sizes, and transition-level PSIS-LOO summaries were retained.

The principal negative B association with μ remained stable across narrow, reference, and wide priors, whereas context coefficients remained close to zero with intervals spanning zero (Supplementary Fig. S3). Because influential observations produced elevated Pareto-k values under several prior regimes, absolute transition-level PSIS-LOO values were treated as diagnostic rather than as definitive measures of prior-regime superiority.

Posterior-predictive calibration audit

Posterior-predictive calibration was evaluated for the corrected original-scale x endpoint using models M_0, M_B, M_E, and M_C. For each transition, replicated values

$$x_j^{\mathrm{rep}}$$

were obtained from the posterior-predictive distribution.

Point-prediction accuracy was summarized using the posterior-predictive median $\widetilde{x}_j^{\mathrm{rep}}$ and residual

$$r_j = x_j - \widetilde{x}_j^{\mathrm{rep}}.$$

Mean absolute error was calculated as

$$\mathrm{MAE} = \frac{1}{N}\sum_{j=1}^{N} |r_j|,$$

and root mean squared error as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} r_j^2}.$$

Calibration was assessed using empirical coverage of nominal 50%, 80%, and 95% posterior-predictive intervals. For nominal interval level $1-\alpha$,

$$\widehat{C}_{1-\alpha} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}\left[x_j \in \left(Q_{\alpha/2,j}, Q_{1-\alpha/2,j}\right)\right],$$

where $Q_{\alpha/2,j}$ and $Q_{1-\alpha/2,j}$ are posterior-predictive quantiles for transition j.

A posterior-predictive probability integral transform (PIT)-type summary was additionally calculated for each observation as the fraction of posterior-predictive draws less than or equal to the observed value. Because x is a semiquantitative coded endpoint rather than a continuously measured biological variable, PIT values were used descriptively rather than subjected to a formal uniformity test.

Calibration was summarized across all transitions and separately by the TNK-dominant and B/Myeloid-shifted candidate ecological contexts. Context-stratified 95% coverage is shown in Supplementary Fig. S5.

Simulation-based parameter recovery

Simulation-based parameter recovery was used to evaluate whether the revised OU model could recover moderate ecological effects under the sparse and irregular sampling structure of the empirical longitudinal cohort.

Rather than simulating trajectories on an artificial regular time grid, each simulated dataset retained the observed participant identities, ecological covariates, numbers of transitions, and participant-specific Δt sequences of the 34-participant x cohort. This design preserves the empirical irregularity and sparsity relevant to parameter identification.

For each participant, simulated trajectories were generated recursively from the exact OU transition density. Conditional on the preceding simulated state $X_i(t)$,

$$X_i(t+\Delta t) \sim \mathcal{N}\left[\mu_i + \big(X_i(t) - \mu_i\big)e^{-\theta_i \Delta t},\ \frac{\sigma^2_{\text{proc}}}{2\theta_i}\left(1 - e^{-2\theta_i \Delta t}\right)\right].$$

The generated state at one transition was used as the preceding state for the next transition from the same participant.

Two generating scenarios were considered. The null scenario used

$$\mu_0 = 0.25, \qquad \theta_0 = 0, \qquad \sigma_{\text{proc}} = 0.30,$$

with all ecological coefficients set to zero.

The moderate-directional scenario used the same baseline parameters with

$$\beta_{\mu,B} = -0.07,$$

$$\beta_{\mu,\text{Myeloid}} = 0.03,$$

$$\beta_{\theta,B} = 0.20,$$

and

$$\beta_{\theta,\text{Myeloid}} = -0.20.$$

Both the baseline M_0 and continuous ecological M_E formulations were evaluated. Five independently simulated datasets were generated for each scenario-model combination.

Recovery fits used the reference prior specification, two chains, 1,200 tuning iterations, and 800 retained posterior draws per chain. For parameter ψ, recovery was summarized using posterior-mean bias,

$$\text{Bias}(\hat{\psi}) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\psi}_r - \psi\right),$$

root mean squared error,

$$\mathrm{RMSE}(\hat{\psi}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\psi}_r - \psi\right)^2},$$

and empirical 95% HDI coverage,

$$\mathrm{Coverage}_{95} = \frac{1}{R}\sum_{r=1}^{R}\mathbb{I}\left[\psi \in \mathrm{HDI}_{95,r}\right],$$

where $R = 5$ simulation replicates.

Divergent transitions were also recorded for every recovery fit. Recovery of ecological effects on μ and θ was evaluated separately because the observed irregular sampling design may contain different amounts of information about effective attractor location and mean-reversion strength.

Results from the moderate-directional scenario are summarized in Supplementary Fig. S4. Ecological effects on μ showed relatively low bias, whereas effects on θ exhibited larger RMSE. Null simulations produced ecological coefficient estimates centered near zero with approximately nominal interval coverage.

Interpretation of OU parameters and ecological associations

Three analytical quantities were intentionally distinguished throughout the revised analysis. First, **immune composition** refers to observed participant-level TNK, B, and myeloid fractions obtained from the common cell-type mapping workflow.

Second, **candidate ecological contexts** refer to descriptive clusters obtained by coarse-graining these continuous compositions. Their bootstrap stability and sensitivity to raw versus CLR representation were evaluated explicitly. These labels are not interpreted as established leukemia subtypes or invariant biological ecotypes.

Third, **OU parameters** are effective parameters of the stochastic model fitted to the semi-quantitative longitudinal x endpoint. The effective attractor μ and mean-reversion strength θ therefore summarize features of the modeled scalar trajectory. They should not be interpreted as literal biological optima or direct measurements of evolutionary fitness.

Accordingly, an association between B or myeloid composition and μ or θ represents statistical modulation of the effective longitudinal model, not evidence that the corresponding immune compartment causally determines leukemia dynamics. Likewise, OU-like mean reversion is not interpreted as direct evidence of stabilizing evolutionary selection.

Software, audit outputs, and reproducibility

Single-cell processing and common cell-type harmonization were implemented in Python using `Scanpy`, `AnnData`, `pandas`, and `NumPy`. Ecological-context discovery and stability analyses used `scikit-learn` and `SciPy`. Bayesian models were implemented in `PyMC` with `PyTensor`, and posterior summaries and PSIS-LOO diagnostics were generated using `ArviZ`.

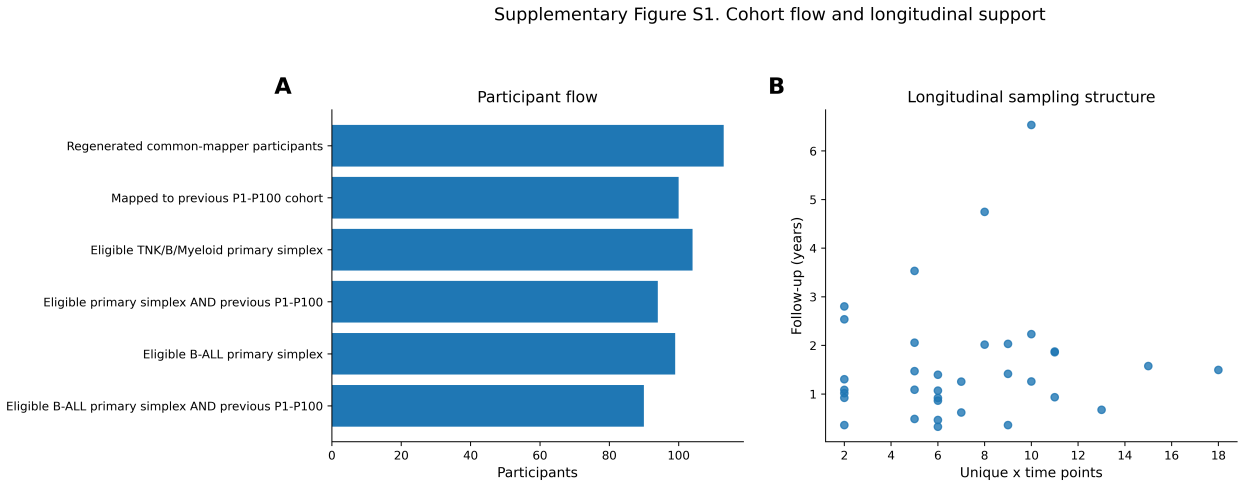
Fixed random seeds were used for K-means clustering, bootstrap resampling, posterior sampling, shuffled-context construction, and simulation experiments. The revised workflow generated explicit audit tables for annotation availability, participant inclusion and exclusion, participant-ID reconciliation, duplicate longitudinal observations, ecological-context assignments, longitudinal support, predictor construction, model diagnostics, posterior summaries, LOPO validation, transformed-x sensitivity, ordinal-state analysis, prior sensitivity, posterior-predictive calibration, and parameter recovery.

All revised main and supplementary figures were regenerated from saved analysis tables and posterior outputs. Figure source tables were exported alongside the final figure-generation workflow to preserve numerical traceability between the statistical analyses and the manuscript graphics.

Scope of the supplementary robustness analyses

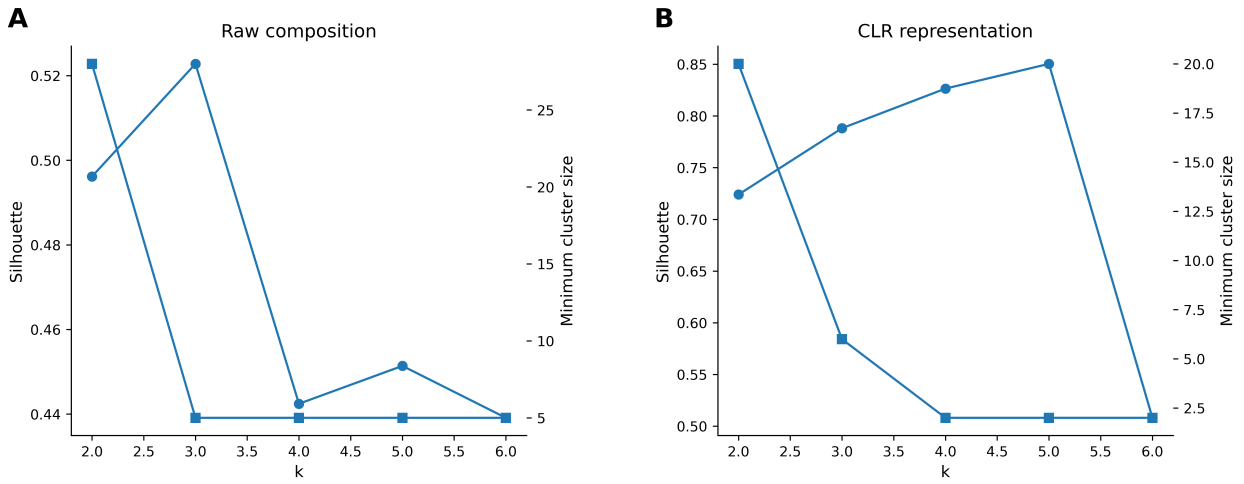
The supplementary analyses were designed to evaluate different potential failure modes rather than to maximize the number of alternative models. Supplementary Fig. S1 documents cohort attrition and longitudinal sampling structure; Supplementary Fig. S2 evaluates cluster-resolution and compositional-representation sensitivity; Supplementary Fig. S3 evaluates prior sensitivity of the principal ecological coefficients; Supplementary Fig. S4 evaluates simulation-based parameter recovery under the observed irregular sampling design; Supplementary Fig. S5 evaluates posterior-predictive calibration separately within the two coarse ecological contexts; and Supplementary Fig. S6 documents influential transitions and robustness of the B-associated attractor coefficient to transformed-x boundary handling. Taken together, these analyses were used to determine which conclusions were reproducible across clustering representation, endpoint scale, prior specification, predictive validation, and simulated recovery. The primary inferential emphasis was placed on continuous ecological composition and participant-level predictive validation rather than on categorical context-specific parameter estimates.

Supplementary Figures

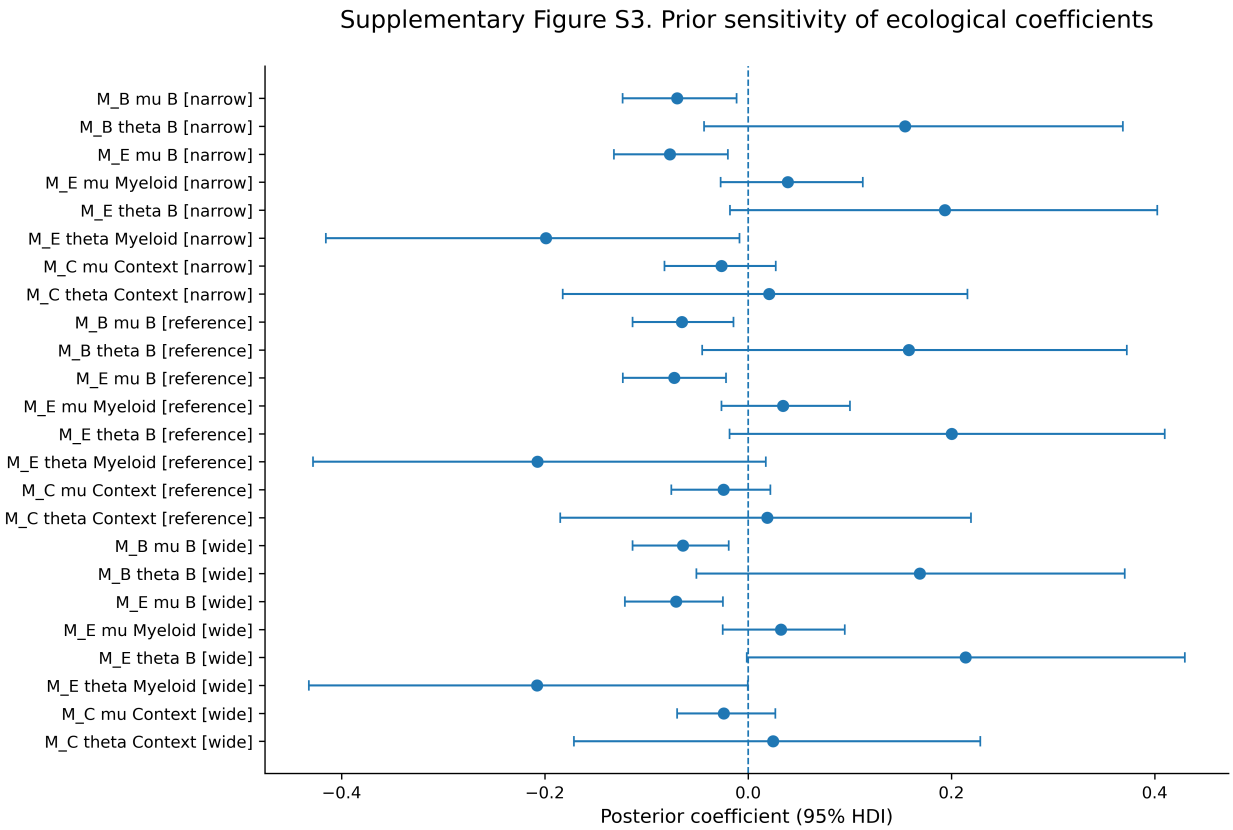


Supplementary Figure S1. Cohort flow and longitudinal support. (**A**) Participant flow through the revised ecological and longitudinal analyses. The common cell-type mapping workflow regenerated immune-compositional profiles for 113 participants; 100 were mapped to the previous P1–P100 cohort, 104 satisfied eligibility criteria for the primary TNK/B/myeloid simplex, and 94 were both compositionally eligible and mapped to the previous cohort. Within B-ALL, 99 participants satisfied primary compositional eligibility and 90 were also mapped to the previous cohort. (**B**) Longitudinal sampling structure for the 34 participants contributing valid transitions of the primary x endpoint. Each point represents one participant and shows the number of unique x observation times versus total follow-up duration. The cohort contained 241 valid x observations and 207 positive-time transitions, with a median of 6 unique observation times per participant (range, 2–18) and median follow-up of 1.283 years. The dispersion illustrates the sparse and irregular longitudinal sampling structure motivating use of the exact irregular-time OU transition likelihood. TNK, combined T-cell/natural-killer compartment; B-ALL, B-cell acute lymphoblastic leukemia; OU, Ornstein–Uhlenbeck.

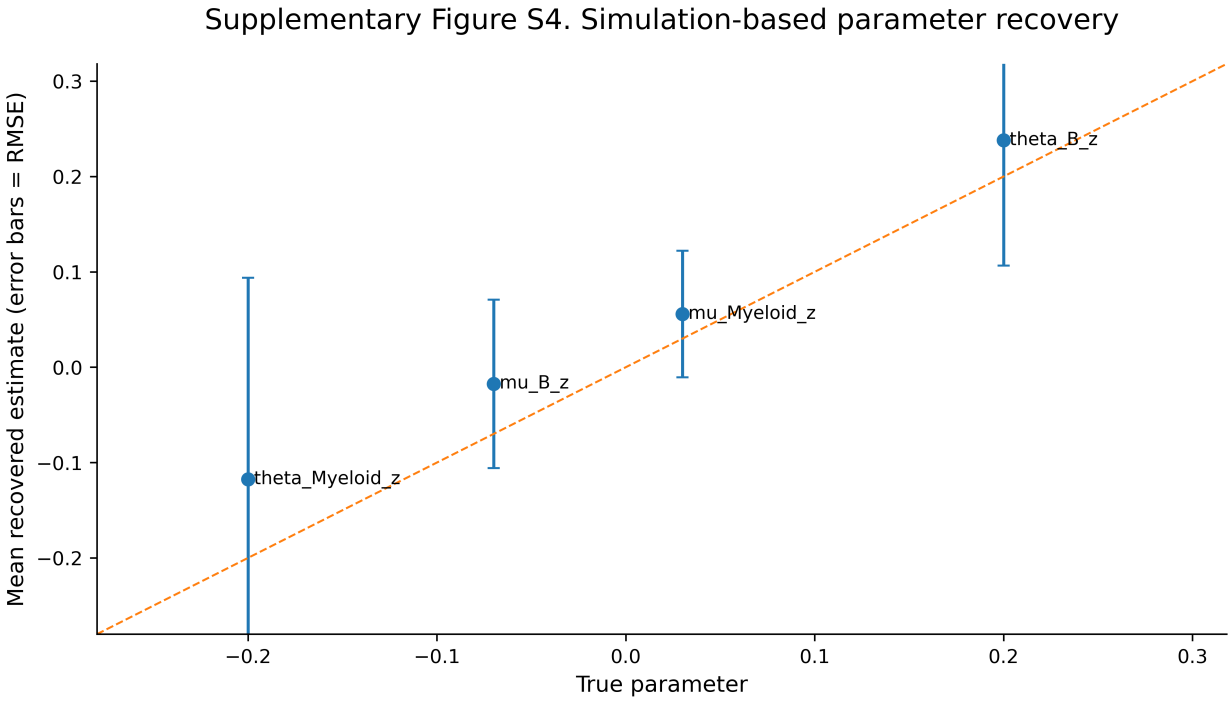
Supplementary Figure S2. Cluster-resolution sensitivity



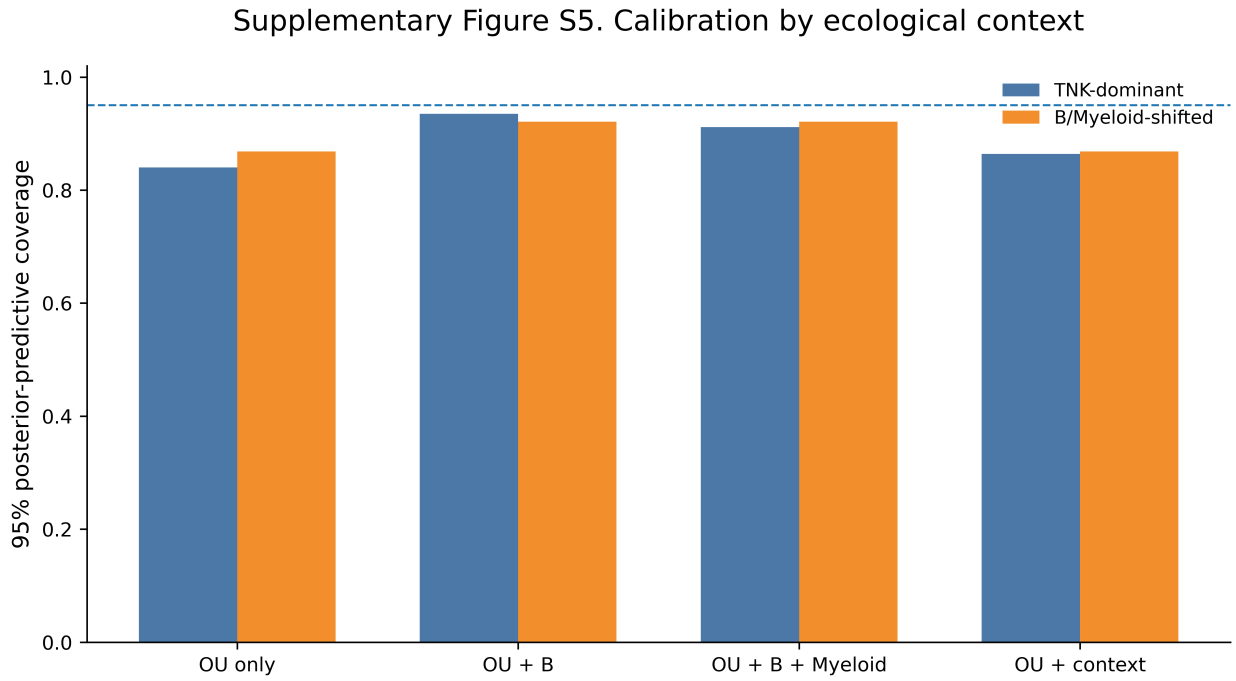
Supplementary Figure S2. Cluster-resolution sensitivity under raw and centered log-ratio representations. (**A**) Sensitivity of K-means clustering to resolution $k = 2$–6 using the raw, renormalized TNK/B/myeloid composition. Circles show silhouette coefficients (left axis), and squares show the minimum number of participants in any cluster at each resolution (right axis). Although separation remains moderate across candidate resolutions, increasing k rapidly produces small clusters, with the minimum cluster size decreasing to five participants from $k = 3$ onward. (**B**) Corresponding analysis using standardized centered log-ratio (CLR) coordinates. CLR-based clustering yields high silhouette coefficients at several resolutions but increasingly imbalanced partitions, including minimum cluster sizes of only two participants at $k \geq 4$. Together with the low raw-versus-CLR adjusted Rand indices shown in main Figure 3D, these results demonstrate that apparently strong within-representation separation does not establish representation-invariant ecological subtypes. CLR transformation used a 10^{-6} pseudocount added to each TNK/B/myeloid component before renormalization and log-ratio transformation.



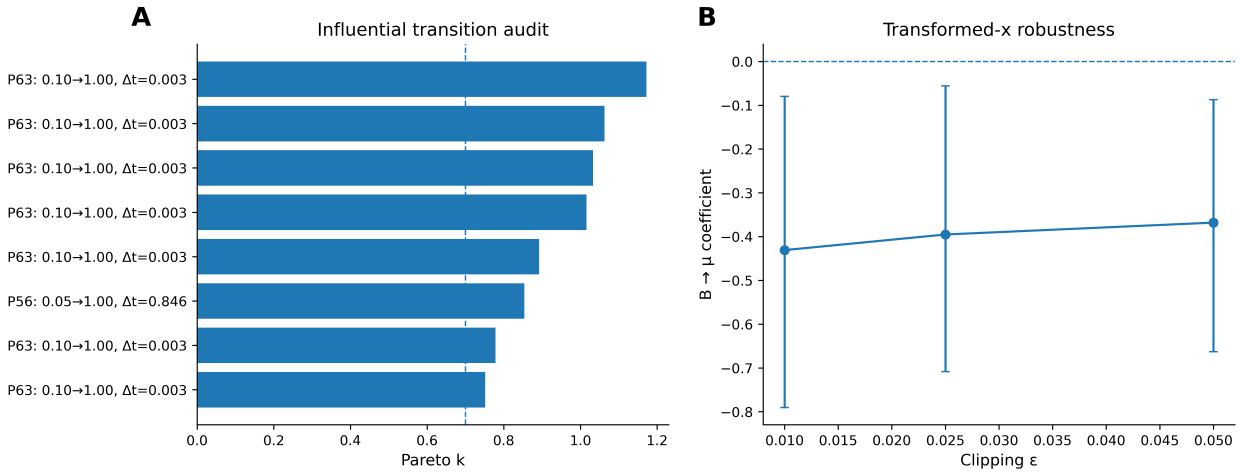
Supplementary Figure S3. Prior sensitivity of ecological coefficients in the revised OU models. Posterior ecological coefficients for the B-only (M_B), B+myeloid (M_E), and context-only (M_C) OU models under narrow, reference, and wide prior regimes. Points denote posterior means and horizontal error bars denote 95% highest-density intervals (HDIs); the vertical dashed line denotes zero. Coefficients modifying the effective attractor μ and log mean-reversion strength θ are displayed separately. The negative B association with μ is stable across all three prior specifications in both M_B and M_E, whereas the coarse ecological-context effect on μ remains close to zero with intervals spanning zero. B- and myeloid-associated effects on θ show consistent directional tendencies but broader uncertainty. These analyses indicate that the principal B-associated attractor shift and the absence of a resolved context effect are not consequences of the reference prior specification. OU, Ornstein–Uhlenbeck; HDI, highest-density interval.



Supplementary Figure S4. Simulation-based parameter recovery under the observed longitudinal sampling design. Recovery of ecological coefficients in the moderate-directional simulation scenario for the B+myeloid OU model (M_E). Simulated trajectories retained the actual participant identities, ecological covariates, numbers of transitions, and irregular Δt sequences of the 34-participant longitudinal cohort. Five independent datasets were generated and refitted. The horizontal axis shows the true generating parameter and the vertical axis shows the mean recovered posterior estimate across simulation replicates; vertical error bars show root mean squared error (RMSE), and the dashed diagonal indicates perfect recovery. The simulated coefficients were $\beta_{\mu,B} = -0.07$, $\beta_{\mu,\text{Myeloid}} = 0.03$, $\beta_{\theta,B} = 0.20$, and $\beta_{\theta,\text{Myeloid}} = -0.20$. Effects on the effective attractor μ were recovered with relatively small bias, whereas effects on mean-reversion strength θ showed greater estimation variability. Complementary null simulations produced ecological coefficients centered near zero with approximately nominal 95% interval coverage, indicating that the revised model did not systematically generate ecological associations when none were simulated. OU, Ornstein–Uhlenbeck; RMSE, root mean squared error.



Supplementary Figure S5. Posterior-predictive calibration by candidate ecological context. Empirical coverage of nominal 95% posterior-predictive intervals for the corrected original-scale x endpoint, stratified by the two coarse candidate ecological contexts. Bars show coverage separately for TNK-dominant and B/Myeloid-shifted transitions under OU-only (M_0), B-only (M_B), B+myeloid (M_E), and context-only (M_C) models; the dashed horizontal line indicates nominal 95% coverage. The B-only model achieved similar coverage in B/Myeloid-shifted (92.1%) and TNK-dominant (93.5%) transitions, while B+myeloid also showed improved coverage relative to OU-only. The similarity between contexts indicates that the calibration improvement associated with continuous ecological covariates is not confined to one coarse ecological group. OU, Ornstein–Uhlenbeck.



Supplementary Figure S6. Influential observations and transformed-x sensitivity. (**A**) Audit of transitions with Pareto-$k > 0.7$ in transition-level PSIS-LOO analyses of the revised OU models. Bars show Pareto-k values, and the dashed vertical line denotes the commonly used $k = 0.7$ diagnostic threshold. The same abrupt P63 transition from $x = 0.10$ to $x = 1.00$ over $\Delta t \approx 0.003$ years is influential across several model specifications, with Pareto-$k > 1$ in multiple models; a P56 transition from $x = 0.05$ to $x = 1.00$ over $\Delta t = 0.846$ years is also influential in the ecology-plus-context model. These observations were retained rather than excluded, and true participant-level refitting was used for the primary predictive comparison instead of relying on transition-level importance sampling. (**B**) Sensitivity of the B-associated effective-attractor coefficient to transformation of the bounded x endpoint. Points show posterior means for the B$\rightarrow \mu$ coefficient from the B-only OU model after logit transformation using clipping values $\epsilon = 0.01$, 0.025, and 0.05; error bars denote 95% HDIs and the dashed horizontal line denotes zero. The B-associated coefficient remains negative across all clipping choices, demonstrating that the principal ecological association is robust to the treatment of upper-bound x values. PSIS-LOO, Pareto-smoothed importance-sampling leave-one-out cross-validation; OU, Ornstein–Uhlenbeck; HDI, highest-density interval.

Supplementary Tables

Supplementary Table S1. Cohort reconciliation, participant inclusion, and longitudinal sampling characteristics.
File:
`Supplementary_Table_S1_cohort_reconciliation_sampling.csv`
This table summarizes participant reconciliation across the two ScPCA projects, mapping to the legacy P1–P100 cohort, eligibility for the primary TNK/B/Myeloid compositional analysis, diagnosis composition, and longitudinal sampling characteristics for participants contributing to the primary longitudinal models.

Supplementary Table S2. Harmonized single-cell annotation mapping and project-level quality-control summary.
Files:

`Supplementary_Table_S2A_project_QC_summary.csv`
`Supplementary_Table_S2B_annotation_mapping_audit.csv`
`Supplementary_Table_S2C_broad_class_summary.csv`

These tables document the common annotation-mapping workflow used to harmonize single-cell annotations across SCPCP000008 and SCPCP000022. They report project-level quality-control metrics, original-annotation-to-broad-class mappings, cell counts, project-level fractions, and aggregate broad-class summaries.

Supplementary Table S3. Candidate ecological-context discovery and stability across clustering specifications.
Files:

`Supplementary_Table_S3A_k_scan_and_raw_CLR_concordance.csv`
`Supplementary_Table_S3B_bootstrap_cluster_stability.csv`
`Supplementary_Table_S3C_BALL_k_scan_and_raw_CLR_concordance.csv`
`Supplementary_Table_S3D_BALL_bootstrap_cluster_stability.csv`

These tables summarize candidate ecological-context resolution and robustness across raw and centered log-ratio (CLR) representations. Reported quantities include cluster sizes, cluster-size coefficient of variation, silhouette coefficients, project and diagnosis Cramér's V, raw–CLR adjusted Rand index, bootstrap global pairwise agreement, and cluster-specific Jaccard stability. B-ALL-restricted analyses provide an additional diagnosis-focused sensitivity assessment.

Supplementary Table S4. Primary longitudinal OU model input and design tables.

Archive:
`Supplementary_Data_4_Primary_Longitudinal_OU_Model_Inputs_and_Design.zip`

Files included in the archive:

`Supplementary_Data_4_manifest.csv`
`Supplementary_Data_4_README.txt`
`wave2_context_summary.csv`
`wave2_model_specifications.csv`
`wave2_n_context_support.csv`
`wave2_n_duplicate_resolution_audit.csv`
`wave2_n_duplicate_time_audit.csv`
`wave2_n_exclusions.csv`
`wave2_n_state_counts.csv`
`wave2_n_transition_matrix_counts.csv`
`wave2_n_transition_matrix_row_proportions.csv`
`wave2_participant_design_matrix.csv`
`wave2_predictor_scaler_parameters.csv`
`wave2_shuffled_context_mapping.csv`
`wave2_x_duplicate_resolution_audit.csv`
`wave2_x_duplicate_time_audit.csv`
`wave2_x_exclusions.csv`
`wave2_x_participant_summary.csv`
`wave2_x_sampling_summary.csv`
`wave2_x_transition_table.csv`

This supplementary archive contains the machine-readable transition-level inputs, participant-level design objects, preprocessing audits, and supporting summaries used to construct the primary longitudinal Ornstein–Uhlenbeck (OU) analysis. The primary continuous-x materials include the canonical longitudinal transition table, participant and sampling summaries, duplicate-time and duplicate-resolution audits, endpoint-specific exclusions, participant design matrix, predictor scaling parameters, model specifications, ecological-context summary, and shuffled-context mapping used for the negative-control analysis. The archive additionally contains context-support, duplicate-resolution, exclusion, state-count, and transition-matrix summaries for the secondary ordinal-n endpoint. A README and file manifest document the contents and analytical role of each archived file. The ordinal-n materials are included as secondary-endpoint robustness outputs and should not be interpreted as substitutes for the primary continuous-x OU input.

Supplementary Table S5. Specification, posterior estimates, diagnostics, and information-criterion comparison of primary OU models.
Files:

`Supplementary_Table_S5A_primary_OU_model_comparison.csv`

This table summarizes the primary OU model family, including M_0, M_B, M_M, M_E, M_C, M_{EC}, and M_S. Reported quantities include active predictors, convergence diagnostics, effective sample sizes, divergences, transition-level PSIS-LOO expected log predictive density, p_{LOO}, and Pareto-k diagnostics. Detailed posterior summaries and ecological coefficient outputs are provided in Supplementary Data 5.

Supplementary Table S6. Participant-level leave-one-participant-out predictive validation of primary OU models.
Files:

`Supplementary_Table_S6A_LOPO_model_summary.csv`
`Supplementary_Table_S6B_LOPO_pairwise_differences.csv`
`Supplementary_Table_S6C_LOPO_context_stratified_summary.csv`
`Supplementary_Table_S6D_MB_vs_M0_participant_results.csv`

These tables report true leave-one-participant-out (LOPO) predictive validation results for the primary OU models. They include model-level held-out expected log predictive density, pairwise participant-level differences, ecological-context-stratified summaries, and participant-wise comparisons between M_B and M_0.

Supplementary Table S7. Robustness analyses: transformed traits, secondary endpoints, and prior sensitivity.
Files:

`Supplementary_Table_S7A_transformed_x_LOPO.csv`
`Supplementary_Table_S7B_secondary_ordinal_coefficients.csv`
`Supplementary_Table_S7C_transformed_x_coefficients_epsilon_sensitivity.csv`
`Supplementary_Table_S7D_prior_sensitivity_model_summary.csv`
`Supplementary_Table_S7E_prior_sensitivity_coefficients.csv`

These tables consolidate robustness analyses involving logit-transformed versions of the primary compositional trait, the secondary ordinal endpoint, transformed-x LOPO validation, and prior sensitivity. They report transformed-trait predictive performance, ecological coefficients, sensitivity to boundary offsets $\epsilon = 0.01$, 0.025, and 0.05, and stability across narrow, reference, and wide prior regimes.

Supplementary Table S8. Parameter recovery and posterior-predictive calibration.
Files:

`Supplementary_Table_S8A1_parameter_recovery.csv`
`Supplementary_Table_S8A2_parameter_recovery_aggregate.csv`
`Supplementary_Table_S8B_PPC_model_summary.csv`
`Supplementary_Table_S8C_PPC_context_summary.csv`

These tables summarize simulation-based parameter recovery and posterior-predictive calibration. Recovery metrics include generating values, posterior mean estimates, bias, root mean squared error, empirical 95% coverage, and divergence counts. Calibration metrics include mean absolute error, root mean squared error, mean residual, posterior-predictive interval coverage, and probability integral transform summaries at the model and ecological-context levels.

Supplementary Table S9. Analysis workflow and provenance of manuscript outputs.
File:
`Supplementary_Table_S9_analysis_workflow_provenance.csv`
This table provides a concise provenance map linking each major analysis stage to the corresponding analysis script, principal inputs, generated outputs, and manuscript figures, supplementary figures, supplementary tables, or supplementary data files.

Supplementary Data

Supplementary Data 1. Harmonized sample- and participant-level TME composition dataset.

File:

`Supplementary_Data_1_Harmonized_TME_Composition.xlsx`

Excel workbook containing sample-level harmonized TME composition for SCPCP000008 and SCPCP000022, the combined sample-level composition matrix, and the combined participant-level composition dataset used for downstream ecological-context analyses.

Supplementary Data 2. Single-cell annotation mapping and audit results.

File:

`Supplementary_Data_2_Single_Cell_Annotation_Audit.xlsx`

Excel workbook containing project-level and sample-level annotation audits for SCPCP000008 and SCPCP000022, original-annotation-to-broad-class mappings, cell counts and fractions, and project-level common-mapper quality-control summaries.

Supplementary Data 3. Participant reconciliation, ecological-context assignments, and clustering sensitivity results.

File:

`Supplementary_Data_3_Reconciliation_Context_Clustering_Sensitivity.xlsx`

Excel workbook containing participant reconciliation, participant-flow summaries, eligible primary composition data, excluded participants, raw and CLR candidate ecological-context assignments for $k = 2$–6, B-ALL sensitivity assignments, clustering-resolution scans, bootstrap stability summaries, and raw–CLR concordance results.

Supplementary Data 4. Primary longitudinal OU model inputs and design tables.

File:

`Supplementary_Data_4_Primary_Longitudinal_OU_Model_Inputs_and_Design.xlsx`

Excel workbook containing the primary continuous-x longitudinal transition table, participant-level design matrix, model specifications, predictor scaling parameters, duplicate-resolution audit, endpoint-specific exclusions, secondary ordinal-n transition data, ecological-context support, state counts, and transition-count and row-proportion matrices. The machine-readable CSV/TXT versions of these materials are provided in the associated Supplementary Data 4 ZIP archive; consolidated tables are additionally provided in the Supplementary Data 4 Excel workbook.

Supplementary Data 5. Primary OU posterior summaries, diagnostics, and full-data model comparison.

File:

`Supplementary_Data_5_Primary_OU_Posterior_Diagnostics_Model_Comparison.xlsx`

Excel workbook containing model-specific and consolidated posterior summaries for M_0, M_B, M_M, M_E, M_C, M_{EC}, and M_S; ecological coefficient summaries; convergence diagnostics; transition-level PSIS-LOO comparisons; complete pointwise Pareto-k results; and influential-transition summaries.

Supplementary Data 6. Leave-one-participant-out predictive validation results.

File:

`Supplementary_Data_6_LOPO_Predictive_Validation.xlsx`

Excel workbook containing model-level true LOPO summaries, pairwise participant-level ELPD differences, complete held-out-participant predictive results, fold-level sampling diagnostics, and participant-wise M_B versus M_0 comparisons.

Supplementary Data 7. Secondary-endpoint, transformed-trait, and prior-sensitivity analyses.

File:

`Supplementary_Data_7_Secondary_Transformed_Prior_Sensitivity.xlsx`

Excel workbook containing secondary ordinal-n coefficient summaries, transformed-x analyses for $\epsilon = 0.01$, 0.025, and 0.05, transformed-x transition tables, transformation audits, transformed-x LOPO validation at $\epsilon = 0.05$, and prior-sensitivity model and coefficient summaries.

Supplementary Data 8. Simulation-based parameter recovery and posterior-predictive calibration results.

File:

`Supplementary_Data_8_Parameter_Recovery_and_PPC.xlsx`

Excel workbook containing parameter-level and replicate-level simulation recovery results together with posterior-predictive calibration summaries at the model, ecological-context, participant, and transition levels. These data underlie Supplementary Figures S4–S5 and Supplementary Table S8.

Supplementary Notes

Supplementary Note 1: Mathematical formulation and interpretation of the revised ecological OU model

The revised longitudinal analysis uses an Ornstein–Uhlenbeck (OU) process as an effective continuous-time stochastic model for the prespecified semiquantitative leukemia-state variable x. For participant i,

$$dX_i(t) = \theta_i\big(\mu_i - X_i(t)\big)\,dt + \sigma_{\text{proc}}\,dW_i(t),$$

where μ_i is the participant-specific effective attractor, $\theta_i > 0$ is the effective mean-reversion strength, $\sigma_{\text{proc}} > 0$ is the shared process-diffusion scale, and $W_i(t)$ is a standard Wiener process.

For consecutive observations separated by interval Δt, the exact OU transition density is

$$X_i(t+\Delta t) \mid X_i(t) \sim \mathcal{N}\left[\mu_i + \big(X_i(t) - \mu_i\big)e^{-\theta_i \Delta t},\ \frac{\sigma_{\text{proc}}^2}{2\theta_i}\left(1 - e^{-2\theta_i \Delta t}\right)\right].$$

This closed-form likelihood permits direct modeling of irregular observation intervals without discretizing time onto a common grid.

In the revised formulation, ecological information enters through regression on the OU parameters rather than through context-specific hierarchical partial pooling. For participant i,

$$\mu_i = \mu_0 + \mathbf{X}_i^{\mathsf{T}}\boldsymbol{\beta}_\mu,$$

and

$$\log \theta_i = \theta_0 + \mathbf{X}_i^{\mathsf{T}}\boldsymbol{\beta}_\theta,$$

so that

$$\theta_i = \exp\left(\theta_0 + \mathbf{X}_i^{\mathsf{T}}\boldsymbol{\beta}_\theta\right).$$

Depending on the ablation model, $\mathbf{X}_i$ contains standardized B composition, myeloid composition, the binary coarse ecological-context indicator, or combinations thereof. Because TNK, B, and myeloid fractions sum to one, B and myeloid are modeled jointly with TNK serving as the implicit compositional reference.

The coefficient vector $\boldsymbol{\beta}_\mu$ describes associations with the effective attractor, whereas $\boldsymbol{\beta}_\theta$ describes associations with log mean-reversion strength. These are statistical parameters

of the effective model and should not be interpreted as direct causal effects of immune compartments on leukemia biology.

The term "attractor" is used in a statistical rather than mechanistic sense. Likewise, OU mean reversion does not establish stabilizing evolutionary selection. The observed x series is a coded semiquantitative leukemia-state proxy rather than a continuously measured molecular or cellular phenotype. Accordingly, the revised analysis interprets μ and θ as effective summaries of longitudinal behavior on the modeled scale.

Supplementary Note 2: Continuous ecological composition versus discrete candidate contexts

The revised analysis distinguishes continuous immune composition from discrete ecological coarse-graining. Participant-level ecological composition is represented by the renormalized TNK, B, and myeloid fractions derived from the common cell-type mapping workflow.

Candidate ecological contexts are obtained by K-means clustering of these participant-level compositions. The primary working resolution is $k = 2$, yielding TNK-dominant and B/Myeloid-shifted groups. These labels are descriptive summaries of relative composition rather than predefined biological states.

Bootstrap resampling showed strong reproducibility of the coarse $k = 2$ structure within the raw-composition representation. However, clustering based on standardized centered log-ratio (CLR) coordinates yielded substantially different participant assignments, with low adjusted Rand index concordance between raw and CLR solutions.

This distinction is central to interpretation. A clustering solution may be internally reproducible under a fixed representation while still failing to define a representation-invariant ecological subtype. Therefore, the revised longitudinal analysis evaluates continuous B and myeloid composition directly and treats the binary candidate context as a separate ablation variable.

The matched model architecture is

$$M_0 : \text{OU only},$$

$$M_B : \text{OU + B},$$

$$M_M : \text{OU + Myeloid},$$

$$M_E : \text{OU + B + Myeloid},$$

$$M_C : \text{OU + context},$$

$$M_{EC} : \text{OU + B + Myeloid + context},$$

and

$$M_S : \text{OU + shuffled context}.$$

This ablation structure separates three questions: whether an OU model captures the longitudinal endpoint; whether continuous ecological composition adds information; and whether coarse context labels add information beyond the continuous variables from which they were derived.

The revised results indicate that continuous B composition carries the clearest ecological association with the effective OU attractor, whereas the coarse context label does not provide a comparable longitudinal contribution. The candidate contexts therefore remain useful for descriptive ecological coarse-graining but are not supported as independently predictive dynamic states in the present cohort.

Supplementary Note 3: Predictive validation, influential observations, and robustness

Transition-level PSIS-LOO was used as a diagnostic measure but not as the primary assessment of predictive generalization because multiple transitions were contributed by the same participant. Leaving out a single transition while retaining other observations from the same participant can underestimate the difficulty of prediction for an unseen participant.

True leave-one-participant-out (LOPO) validation was therefore used as the principal predictive analysis. For each fold, all transitions from one participant were excluded, ecological predictors were standardized using the training participants only, the model was refitted, and the full held-out trajectory was evaluated.

Participant-level predictive differences were calculated as paired ELPD contrasts,

$$\Delta_i = \text{ELPD}_{i,A} - \text{ELPD}_{i,B}.$$

This approach permits direct evaluation of whether an aggregate model advantage is broadly shared or driven by a subset of participants.

An influential-observation audit identified one abrupt P63 transition from $x = 0.10$ to $x = 1.00$ over approximately $\Delta t = 0.003$ years as the primary source of elevated Pareto-k values in several transition-level PSIS-LOO analyses. This observation was retained rather than

removed. Its influence illustrates the mismatch that can occur between a locally smooth Gaussian OU process and abrupt changes in a semiquantitative coded state. Because LOPO relies on explicit refitting rather than importance-sampling approximation, the principal predictive conclusions do not depend on stable PSIS weights for this transition.

Robustness was evaluated through several complementary analyses. First, the bounded x endpoint was logit-transformed after clipping at $\epsilon = 0.01$, 0.025, and 0.05. The negative B-associated effect on μ persisted across all clipping values. Second, true LOPO on the transformed endpoint at $\epsilon = 0.05$ reproduced the qualitative model ranking observed on the original scale. Third, narrow, reference, and wide prior regimes yielded the same qualitative ecological conclusions. Fourth, the longitudinal n series was modeled separately as a three-level ordinal endpoint, avoiding inappropriate pooling of the two longitudinal variables. Finally, simulation-based parameter recovery using the observed irregular sampling schedules showed that moderate ecological effects on μ could be recovered with relatively small bias, while effects on θ were less precisely identified.

These analyses support a conservative interpretation: continuous ecological composition contains modest but reproducible longitudinal information, whereas discrete context membership does not show robust incremental predictive value.

Supplementary Note 4: Posterior-predictive calibration and uncertainty interpretation

Posterior-predictive calibration was evaluated separately from model ranking because a model can provide similar average predictive density while differing in the calibration of predictive uncertainty.

For each transition, posterior-predictive intervals were obtained from the full posterior-predictive distribution. Empirical coverage was evaluated for nominal 50%, 80%, and 95% intervals. For level $1 - \alpha$,

$$\widehat{C}_{1-\alpha} = \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\left[x_j \in \left(Q_{\alpha/2,j}, Q_{1-\alpha/2,j}\right)\right].$$

Point-prediction error was summarized using mean absolute error and root mean squared error, while mean residual was used to assess directional bias.

The continuous ecological models, particularly M_B and M_E, showed modestly lower prediction error and improved 95% predictive coverage relative to OU-only. Calibration was also similar between TNK-dominant and B/Myeloid-shifted transitions, indicating that the coverage improvement was not restricted to one coarse context.

These calibration results complement the participant-level LOPO analysis. The continuous

ecological covariates do not produce a large or uniform gain in held-out predictive density, but they can improve the uncertainty envelope around predictions. For sparse longitudinal data, this distinction is important because uncertainty calibration may be scientifically useful even when point-prediction gains are small.

Supplementary Note 5: Scope and relation to broader OU–branching models

The present analysis uses an OU process only and does not infer explicit branching events, split times, clone birth processes, lineage trees, or branch-specific state trajectories from the ScPCA data.

The broader OU–branching framework remains conceptually relevant for leukemia evolution because longitudinal tumor dynamics may involve both constrained continuous variation and lineage diversification. However, the data used here provide ecological composition at the participant level and a sparse semiquantitative longitudinal state series rather than directly observed lineage-resolved clonal trajectories.

Accordingly, the current model addresses a narrower question: whether single-cell-derived ecological composition contributes interpretable information to effective continuous-time longitudinal dynamics. The revised findings support a modest role for continuous ecological covariates, particularly B composition, but do not support treating the candidate ecological contexts as branch-defining states or stable leukemia subtypes.

Future extensions could combine repeated single-cell or spatial measurements, explicit malignant-cell annotation, molecular clone tracking, and lineage-resolved genomic data within a joint ecological OU–branching model. Such data would permit direct testing of whether ecological composition is associated with attractor shifts, altered mean-reversion rates, branch initiation, or post-branch expansion.

Supplementary Note 6: Summary of inferential scope

The revised analysis supports the following hierarchy of conclusions.

1. Harmonized TNK/B/myeloid composition shows reproducible coarse structure across the analyzed ScPCA projects.

2. The $k = 2$ raw-composition partition is bootstrap-stable and minimally associated with project at coarse resolution, including within B-ALL.

3. Discrete cluster assignments are sensitive to compositional representation, as demonstrated by low raw-versus-CLR concordance.

4. Continuous B composition shows the clearest association with the effective OU attractor μ.

5. The coarse ecological-context label shows little independent association with OU parameters and does not improve participant-level predictive performance.

6. The aggregate predictive advantage of the B-only model is modest and heterogeneous across participants.

7. The B-associated result is qualitatively robust to endpoint transformation and prior specification and is directionally consistent in the separate ordinal n analysis.

8. Simulation-based recovery supports estimation of moderate ecological effects on μ under the observed sampling design, while effects on θ are less precisely identified.

9. Posterior-predictive calibration improves modestly when continuous ecological composition is included.

10. None of these results establishes fixed leukemia ecotypes, direct causal microenvironmental effects, stabilizing evolutionary selection, or clinically validated prediction.

The principal methodological conclusion is therefore that ecological clustering and longitudinal dynamic relevance should be evaluated separately. Stable cross-sectional ecological structure does not necessarily imply a predictive longitudinal state, and continuous ecological covariates may retain information that is lost when the same composition is discretized into cluster labels.

Supplementary References

[1] Uhlenbeck GE, Ornstein LS. On the theory of the Brownian motion. *Physical Review*. 1930;36(5):823–841.

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

[3] Butler MA, King AA. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*. 2004;164(6):683–695.

[4] Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*. 2014;15:1593–1623.

[5] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: CRC Press; 2013.

[6] Abril-Pla O, Andreani V, Carroll C, Dong L, Fonnesbeck CJ, Kochurov M, Kumar R, Lao J, Luhmann CC, Martin OA, Osthege M, Vieira R, Wiecki TV, Zinkov R. PyMC: a modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*. 2023;9:e1516.

[7] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. 2017;27:1413–1432.