Relational Anatomy of Cell Fate Across Lineage-Resolved Perturbation Experiments

A finite relational calculus for state, identity, history, and fate across biological systems

Zed James

Preprint – September 25, 2026.
doi:10.5281/zenodo.22969066.

Abstract

Lineage-resolved single-cell experiments jointly measure molecular state, ancestry, temporal history, and terminal phenotype. We develop a finite relational calculus and audit framework that distinguishes fate enrichment, exact sufficiency, lineage fingerprinting, shared compression, repeated realization, and molecular realization on explicit finite carriers.

Across ReSisTrace, Rewind, and Watermelon, measured precursor states refine later outcome partitions while retaining explicit mixed-fate fibers. Watermelon histories can recover terminal fate exactly after becoming lineage-identifying, and fate-blind compression exposes the boundary between individual fingerprinting and a representation shared across lineages. Repeated day-14 realization generates a support-indexed terminal fate hierarchy with a nonrandom whole-profile geometry and a heterogeneous terminal molecular repertoire.

The same calculus is stress-tested in independent CellTag-multi hematopoietic and reprogramming carriers using RNA, chromatin accessibility, multimodal state, temporal history, technical fate arms, and biological replicate histories. Across 849 canonical state partitions, exact recovery overwhelmingly coincides with lineage-identifying representations; none of the registered noninjective exact candidates survives the complete control stack. At the same time, fate-selective future resolution beyond generic lineage-identity resolution is control-qualified in two system groups.

Across the registered representation class, these results identify an empirical identity boundary for exact state–fate recovery while retaining measurable future alignment below exactness. The resulting cross-carrier geometry supports constrained future realization: measured biological information can resolve future discordance more strongly than generic lineage-identity resolution without generally collapsing future ambiguity to a shared exact fate law.

1 Introduction

Single-cell transcriptomics resolves heterogeneous molecular states at high dimensionality. Lineage tracing supplies ancestry information across destructive molecular measurements. Their combination has made state-to-fate relationships experimentally accessible across development, reprogramming, cancer evolution, and therapeutic response [2, 3, 10, 12, 13, 32]. Barcode-based and CRISPR-based lineage methods provide direct finite ancestry relations among observed cells [5, 6, 4, 7]; trajectory and fate-inference methods provide complementary state-based or state-plus-direction estimates of future behavior [9, 11, 16, 17, 18, 19].

Cancer therapy provides an experimentally useful setting for these questions. Non-genetic heterogeneity can precede drug resistance, rare transcriptional states can enrich for later resistant behavior, and drug-tolerant persister populations can occupy multiple proliferative and metabolic programs [25, 20, 26, 21, 23]. ReSisTrace directly uses sister-cell information to reconstruct pre-treatment states associated with later resistance to olaparib, carboplatin, and natural-killer-cell challenge [22]. Rewind links drug-naive barcode twins to later vemurafenib-resistant clones and resolves substructure within rare precursor states [21]. Watermelon simultaneously records clonal identity, proliferative state, and transcriptome across osimertinib exposure, identifying cycling and non-cycling persister lineages with distinct programs [23]. These experiments provide measured ancestry and state information at a resolution unavailable to classical population-level resistance assays such as the FGF2/imatinib K562 system [24].

A complementary authority is supplied by CellTag-multi, which captures clonal lineage across scRNA-seq and scATAC-seq and applies the same lineage logic to mouse hematopoiesis and fibroblast-to-endoderm-progenitor reprogramming [34]. These systems expand the state coordinate from transcript abundance to orthogonal chromatin accessibility and provide a second temporal program with two biological replicates. They therefore permit an independent test of whether additional modality and temporal depth reveal a shared fate-sufficient state, or instead move the measured representation toward clone identity.

A central analytical distinction follows from the finite structure of these experiments. Association, classification accuracy, lineage enrichment, exact recovery, history sufficiency, and replicate stability are different mathematical statements. Let $\mathcal{L}$ be a finite lineage carrier, let

$$P : \mathcal{L} \longrightarrow \mathcal{S} \tag{1}$$

be a measured state map, and let

$$F : \mathcal{L} \longrightarrow \mathcal{F} \tag{2}$$

be a registered terminal fate map. Exact fate recovery is the fiber-constancy condition

$$\boxed{P(\ell_1) = P(\ell_2) \Longrightarrow F(\ell_1) = F(\ell_2).} \tag{3}$$

Equivalently, a map $\widetilde{F} : \operatorname{im} P \to \mathcal{F}$ exists such that

$$F = \widetilde{F} \circ P. \tag{4}$$

A single pair

$$P(\ell_1) = P(\ell_2), \qquad F(\ell_1) \neq F(\ell_2) \tag{5}$$

provides a complete finite nonrecovery witness.

Throughout this paper, *exact finite recovery* means exact factorization on the observed finite carrier under the specified state and fate representations:

$$\boxed{\text{exact finite recovery} = \text{carrier-scoped factorization under fixed representations.}} \tag{6}$$

It therefore identifies a property of the measured carrier and representation. Claims about biological determinism beyond that carrier require additional experimental authority.

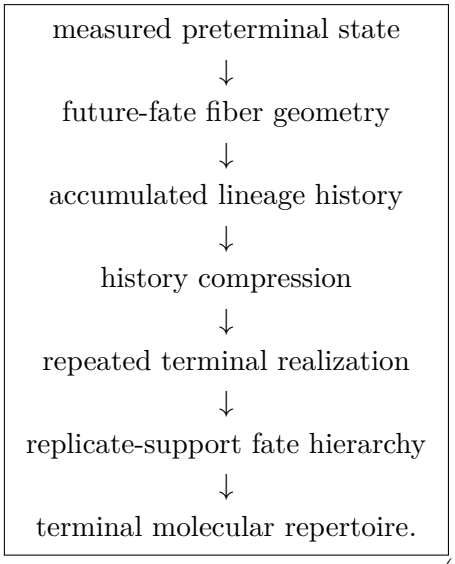
The term *registered* denotes an analysis specification committed and hash-frozen before the corresponding decision computation. Several tranche-level analysis plans were prospectively frozen before outcome evaluation within that analysis stage. This manuscript does not use *preregistered* to imply an external public preregistration of the complete research program.

The analysis defines a *finite relational calculus* for lineage-resolved state–fate experiments. Its primitive objects are finite lineage carriers, measured state maps, temporal histories, terminal fate maps, biological replicate realizations, and molecular-support sets. Its operations are fiber testing, history accumulation, fate-blind quotienting, replicate-relation closure, support filtration, and molecular-support intersection. Each operation answers a distinct inferential question and determines which subsequent claims are warranted. Prediction accuracy alone does not distinguish a representation that captures shared fate structure from one that uniquely identifies each observed lineage. The calculus therefore separates association from sufficiency, sufficiency from lineage fingerprinting, fingerprinting from shared compression, single realization from persistent fate identity, and persistent fate identity from molecular realization. The calculus is the formal system; the audit is its application to an experiment.

The contribution is not new foundational mathematics; it is a domain-specific calculus that assigns explicit biological semantics and inferential roles to finite relational operations. The novelty lies in how these operations are composed to distinguish claims that are often collapsed under the single language of prediction.

Because these operations are finite and explicit, their core relational propositions are machine checked in Lean 4 independently of the empirical carrier constants. The empirical results are reproduced separately from paper-specific finite carriers and deterministic validation artifacts. A public reproducibility repository at `https://github.com/zedjames/relational-cell-fate` contains the released empirical result tables, source provenance, paper-specific validation code, and figure-generation workflow. The broader private development environment and general-purpose formal calculus are not part of that public release.

This paper applies that criterion across successive increases in measurement authority. Pre-treatment state, accumulated temporal history, compressed history, repeated terminal realization, persistent replicate support, and terminal molecular repertoire are retained as separate carriers. The resulting analysis is summarized by progressive and descending structure seen on the right.



(7)

The main results form one continuous state–identity–future geometry. First, measured pre-treatment partitions in ReSisTrace and Rewind retain mixed future-fate fibers under increasing state resolution. Second, Watermelon snapshots retain mixed distributional-fate fibers, while accumulated history reaches exact recovery only at or near lineage-identifying resolution; fate-blind compression exposes where pooled exactness is lost. Third, repeated terminal realization produces a support-indexed fate hierarchy whose whole-profile geometry is strongly structured and whose terminal molecular realization remains heterogeneous. Fourth, orthogonal RNA–ATAC state and temporal multimodal history in CellTag carriers reduce fate collisions further, while exact fine-fate recovery again accumulates at clone-identifying representations. Finally, the complete registered surface bank resolves the distinction quantitatively: exact recovery concentrates at the identity boundary, whereas control-qualified fate-selective alignment remains measurable below exactness.

The paper is scoped to registered finite carriers, measured coordinates, and frozen future definitions in the cited experiments. The conclusions concern exact recoverability, identity resolution, quotient geometry, repeated realization, persistence, modality, temporal history, cross-carrier future alignment, and terminal molecular localization. Causal mechanism and universal state–fate laws remain subsequent experimental questions.

Finite relational calculus for lineage-resolved state–fate inference

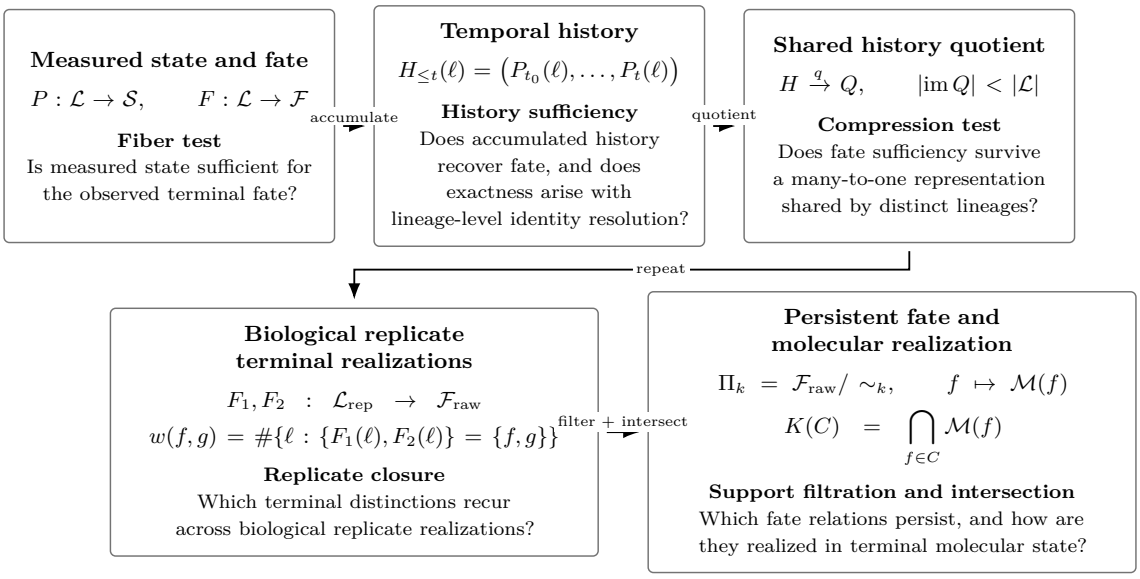


Figure 1: The calculus treats measured state, temporal history, shared quotient structure, biological replicate terminal realization, persistent fate, and molecular support as distinct finite objects. Fiber testing, history accumulation, fate-blind quotienting, replicate closure, support filtration, and molecular-support intersection answer different inferential questions and determine which biological claims are warranted at each stage.

2 Finite state–fate geometry

2.1 Recovery and mixed fibers

For a finite carrier $\mathcal{L}$, define the state fiber of $s \in \operatorname{im} P$ by

$$P^{-1}(s) = \{\ell \in \mathcal{L} : P(\ell) = s\}. \tag{8}$$

Its terminal image is

$$F(P^{-1}(s)) = \{F(\ell) : P(\ell) = s\}. \tag{9}$$

Recovery is equivalent to

$$\boxed{\forall s \in \operatorname{im} P, \qquad \left|F(P^{-1}(s))\right| \le 1.} \tag{10}$$

The mixed-fiber set is

$$\mathcal{A}(P,F) = \left\{ s \in \operatorname{im} P : \left|F(P^{-1}(s))\right| > 1 \right\}. \tag{11}$$

A finite majority residual is

$$R(P,F) = \sum_{s \in \operatorname{im} P} \left(\left|P^{-1}(s)\right| - \max_{f \in \mathcal{F}} \left|P^{-1}(s) \cap F^{-1}(f)\right| \right). \tag{12}$$

Thus

$$\text{Recovers}(P,F) \quad \Longleftrightarrow \quad R(P,F) = 0. \tag{13}$$

2.2 History and compression

For measured times $t_0 < \cdots < t_j$, define a history coordinate

$$H_{\le t_j}(\ell) = \Big(P_{t_0}(\ell), \ldots, P_{t_j}(\ell)\Big) \tag{14}$$

on the lineages with the registered time support. A fate-sufficient history obeys

$$\mathrm{Recovers}(H_{\le t_j}, F). \tag{15}$$

Injectivity provides an automatic finite factorization:

$$H_{\le t_j}(\ell_1) = H_{\le t_j}(\ell_2) \Longrightarrow \ell_1 = \ell_2 \Longrightarrow F(\ell_1) = F(\ell_2). \tag{16}$$

A shared biological quotient therefore requires a noninjective compression

$$q : \mathrm{im}\, H \longrightarrow \mathcal{Q}, \qquad Q = q \circ H, \tag{17}$$

with

$$|\mathrm{im}\, Q| < |\mathcal{L}| \tag{18}$$

and

$$\mathrm{Recovers}(Q, F). \tag{19}$$

These two conditions distinguish exact finite fate sufficiency from lineage fingerprinting.

2.3 Repeated realization and persistent fate

For lineages observed in two terminal replicates, write

$$F_1, F_2 : \mathcal{L}_{\mathrm{rep}} \longrightarrow \mathcal{F}_{\mathrm{raw}}. \tag{20}$$

The direct replicate relation joins $f, g \in \mathcal{F}_{\mathrm{raw}}$ when a lineage realizes f in one replicate and g in the other. Its support is

$$w(f, g) = \#\left\{\ell \in \mathcal{L}_{\mathrm{rep}} : \{F_1(\ell), F_2(\ell)\} = \{f, g\}\right\}. \tag{21}$$

For every support threshold $k \ge 1$, define

$$f \, R_k \, g \quad \Longleftrightarrow \quad f \neq g \text{ and } w(f, g) \ge k, \tag{22}$$

and let $\sim_k$ be its equivalence closure. The resulting partitions satisfy

$$\Pi_{k+1} \text{ refines } \Pi_k. \tag{23}$$

Pairwise persistence is the maximum bottleneck support

$$\lambda(f, g) = \max_{\gamma : f \rightsquigarrow g} \min_{e \in \gamma} w(e). \tag{24}$$

This construction retains the complete graded terminal relation across all registered support thresholds.

2.4 Terminal molecular support

For a raw fate f, let

$$\mathcal{M}(f) \subseteq \mathcal{M}_{\text{state}} \tag{25}$$

be the observed finite support of fate-blind terminal molecular states. For a persistent fate component $C \subseteq \mathcal{F}_{\text{raw}}$, define

$$K(C) = \bigcap_{f \in C} \mathcal{M}(f). \tag{26}$$

and the raw-fate-exclusive kernel

$$K^{\text{ex}}(C) = \{m \in K(C) : \{f : m \in \mathcal{M}(f)\} \subseteq C\}. \tag{27}$$

The molecular analysis uses these exact support objects together with lineage-weighted distributional overlap and matched null geometries.

3 Measurement authority and preterminal state

3.1 Authority-limited branch provenance in FGF2-mediated CML resistance

The FGF2/imatinib K562 system establishes a useful measurement boundary [24]. The source contains bulk parental state, acute treatment/rescue measurements, terminal resistant cultures, and limited intermediate observations. The lineage-linked prebranch interval connecting a measured precursor state to one of the terminal culture identities is absent from the public experimental record. The registered chronology is therefore

$$\underbrace{S_0}_{\text{measured}} \longrightarrow \underbrace{S_1, S_2}_{\text{measured}} \longrightarrow \underbrace{\qquad}_{\text{unmeasured lineage interval}} \longrightarrow \underbrace{K_1, \ldots, K_5}_{\text{measured terminal cultures}}. \tag{28}$$

The available prebranch biology cannot support an exact lineage-specific recovery test. This authority boundary supplies the criterion for the subsequent datasets: a valid preterminal fate experiment requires a measured early state and an explicit lineage bridge to a positively observed later outcome.

3.2 Measured pre-treatment state in ReSisTrace

ReSisTrace satisfies that criterion with a sister-cell design [22]. The reconstructed carrier contains

$$\text{37,011 pre-treatment cells,} \qquad \text{40,713 post-treatment cells,} \tag{29}$$

with

$$\text{2,186 linked lineages,} \qquad \text{6,926 explicit pre/post sister-bridge pairs.} \tag{30}$$

The analytical pre-treatment carrier contains 32,642 cells. The terminal observation distinguishes lineages with a post-treatment surviving sister from lineages lacking a registered surviving sister. The second category retains sampling and detection ambiguity, so the exact result is stated at the observation-disposition level.

Let $P_i^{(c,r)}$ denote the registered pre-treatment state package i in condition c and replicate r, and let $O^{(c,r)}$ denote the later observation disposition. Across carboplatin, olaparib, natural-killer-cell challenge, and control, each with two replicates, every decision-eligible package has at least one witness

$$P_i^{(c,r)}(x) = P_i^{(c,r)}(y), \qquad O^{(c,r)}(x) \neq O^{(c,r)}(y). \tag{31}$$

Hence

$$\boxed{\text{Recovers}(P_i^{(c,r)}, O^{(c,r)}) = \text{false}} \tag{32}$$

for every decision-eligible state package and every condition-by-replicate stratum.

A richer quality-control-augmented state reduces the raw residual in both replicates of all four conditions. The registered fate-permutation controls place the observed refinement within the corresponding null distribution. This preserves a clean distinction between finer molecular partitioning and fate information.

3.3 Positive-fate nonrecovery in Rewind

Rewind supplies a stronger terminal authority because both compared terminal groups are positively realized resistant outcomes [20, 21]. The analyzed drug-naive Carbon Copy carrier contains 297 cells, of which 162 belong to the positively resistant-fate precursor surface. The registered positive-fate split contains

$$77 \text{ cells in the top-30 resistant-barcode abundance pool}, \qquad 85 \text{ cells in the next-30 pool}. \tag{33}$$

Nine source-native RNA-FISH genes define the measured precursor coordinate bank. Thirteen decision-eligible packages range from constant primed state through individual-gene, multigene, and richer finite precursor partitions.

For every package P_i, at least one pair satisfies

$$P_i(x) = P_i(y), \qquad F_R(x) \neq F_R(y), \tag{34}$$

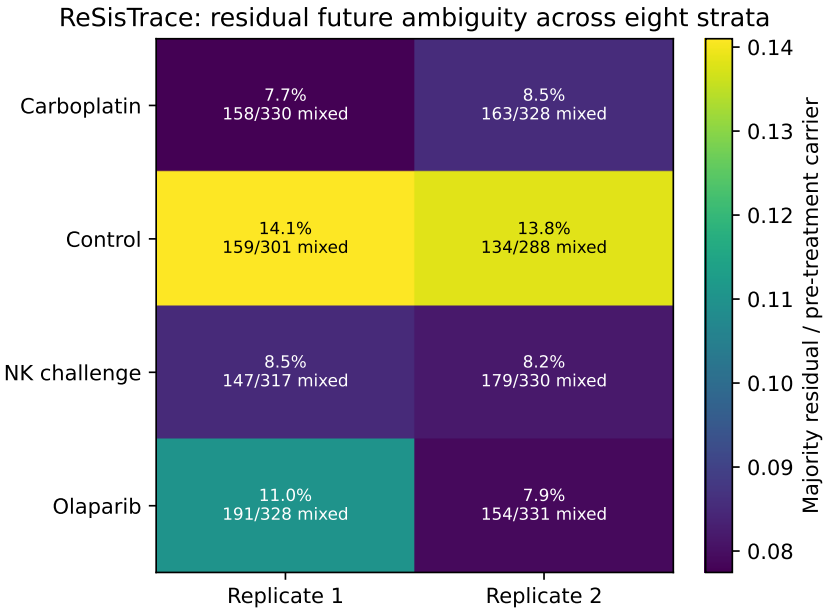
where F_R is the positively realized resistant-fate class. Therefore

$$\boxed{\forall i \in \{1, \ldots, 13\}, \qquad \text{Recovers}(P_i, F_R) = \text{false}.} \tag{35}$$

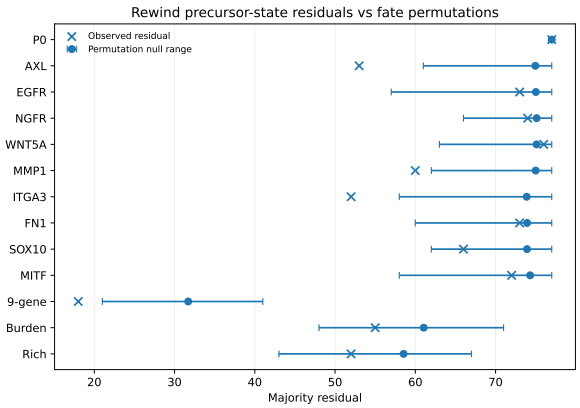
For the joint nine-gene state, 162 cells occupy 86 precursor-state fibers. Seventy-one fibers are fate-pure and 15 are mixed; 63 cells lie in the mixed fibers. The majority residual decreases from 77 under the constant precursor state to 18 under the nine-gene state. Among 1,000 fate permutations, none reaches a residual as small, giving the finite corrected tail

$$p_{\text{perm}} = \frac{1}{1001}. \tag{36}$$

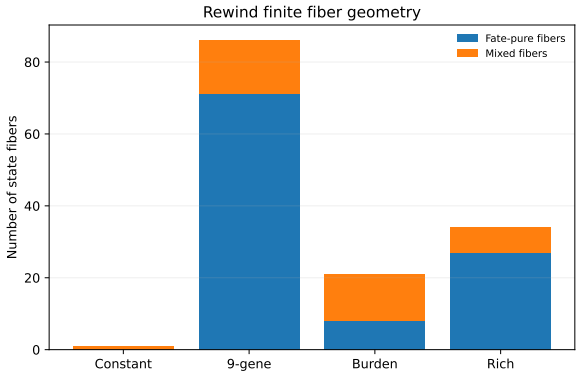
The source-highlighted gene set was identified in relation to the original future behavior, so this result is interpreted as strong measured contraction of future ambiguity and exact positive-fate nonrecovery at the registered pooled-probe resolution.



(a) ReSisTrace P5b corrected-state geometry across four perturbations and two biological replicates. Cell color reports majority residual; annotations give residual percentage and mixed fibers among occupied fibers. All eight strata retain explicit future ambiguity.



(b) Rewind majority residuals for decision-eligible precursor-state representations against 1,000 fate permutations. The joint nine-gene representation produces the smallest registered residual.



(c) Finite Rewind precursor-state fiber geometry. The joint nine-gene state partitions 162 positive-fate precursor cells into 86 occupied fibers: 71 fate-pure and 15 mixed, with 63 cells in mixed fibers.

Figure 2: **Preterminal state and finite future-fate geometry across two lineage-resolved systems.** ReSisTrace uses sister-linked pre-treatment measurements to evaluate future observation disposition across eight condition–replicate strata. Rewind evaluates positively realized resistant-fate classes from measured drug-naive precursor states. In both systems, increased state resolution refines the observed future partition while preserving explicit mixed-fiber structure. Exact finite recovery is therefore evaluated from the occupied state fibers rather than inferred from enrichment or classification performance.

4 Temporal state, history, and finite fingerprinting

4.1 Watermelon temporal carrier

Watermelon records clonal origin, proliferative state, and transcriptome during osimertinib exposure [23]. The reconstructed PC9 carrier contains

$$56{,}419 \text{ measured cells}, \qquad 35{,}792 \text{ barcode-positive cells}, \qquad 8{,}847 \text{ canonical barcode-set lineages}. \tag{37}$$

At day 14,

$$4{,}802 \tag{38}$$

lineages are positively observed. Exactly

$$748 \tag{39}$$

lineages are observed at days 0, 3, 7, and 14.

Terminal fate retains the complete observed distribution among cycling (C), moderately cycling (M), and non-cycling (N) cells. For lineage ℓ,

$$F_{14}(\ell) = \mathrm{reduce}_{\mathrm{gcd}}\left(n_C(\ell), n_M(\ell), n_N(\ell)\right). \tag{40}$$

The 4,802 terminal lineages occupy

$$225 \tag{41}$$

distinct distributional fate objects. The terminal carrier includes pure and mixed proliferative outcomes, including lineages represented in multiple terminal sorting classes.

The reduction in Eq. (40) encodes terminal composition independently of clone size. Thus 1:1:0 and 50:50:0 share one composition object, while 2:1:0 and 3:1:0 remain distinct. This representation is a registered finite terminal coordinate, rather than a latent biological "true fate." Repeated terminal realization and count-qualified sensitivity analyses directly audit its sampling dependence. Requiring at least $1, 2, 3$, or 5 terminal cells in each replicate yields paired carriers of $1550, 654, 367$, and 178 lineages, with $190, 188, 181$, and 156 raw fate objects and maximum nonself supports $110, 30, 13$, and 6. Because support is a lineage count, the corresponding carrier-normalized supports are $7.10\%, 4.59\%, 3.54\%$, and 3.37% (Fig. 5). The count-qualified analysis is therefore reported in both raw and normalized units.

A second revision-stage robustness surface replaces exact GCD-reduced composition by a coarser four-part simplex representation. Each observed (C, M, N) count vector is normalized and assigned, by deterministic largest-remainder apportionment, to an integer composition (a, b, c) satisfying $a + b + c = 4$. This produces 15 possible terminal composition states. The resulting persistence hierarchy remains strongly structured under 1,000 marginal-preserving replicate-pair permutations: class-resolution area, pair-discrimination area, and lineage-persistence area each have corrected upper-tail probability 1/1001 (Fig. 6). This alternative representation is a robustness analysis and does not replace the registered GCD-reduced primary endpoint.

4.2 Instantaneous nonrecovery

A common fate-blind state architecture is evaluated at days 0, 3, and 7. The richer P4 state retains twelve deterministic transcriptomic projections. Its majority residuals are

$$R_{P4}(0) = 10, \qquad R_{P4}(3) = 53, \qquad R_{P4}(7) = 17. \tag{42}$$

A richer distributional P5 state gives

$$R_{P5}(0) = 1, \qquad R_{P5}(3) = 21, \qquad R_{P5}(7) = 6. \tag{43}$$

Thus

$$\boxed{\text{Recovers}(P_t, F_{14}) = \text{false}, \qquad t \in \{0, 3, 7\},} \tag{44}$$

for every decision-eligible instantaneous state package. Both day-3 replicates and both day-7 replicates retain mixed fibers independently.

The residual sequence is nonmonotone. Measured time therefore adds biological information through a sequence of distinct instantaneous partitions.

4.3 History recovery and identity resolution

Accumulated history changes the factorization geometry. For example,

$$H_{0,3}^{(P2)}(\ell) = \left(P_0^{(P2)}(\ell), P_3^{(P2)}(\ell)\right) \tag{45}$$

occupies 1,183 distinct history states on 1,183 registered lineages. Likewise,

$$H_{0,3,7}^{(P2)} \tag{46}$$

occupies 748 states on the 748 complete trajectories. Exact fate factorization follows:

$$\text{Recovers}(H_{0,3}^{(P2)}, F_{14}) = \text{true}, \qquad \text{Recovers}(H_{0,3,7}^{(P2)}, F_{14}) = \text{true}. \tag{47}$$

The same identity-resolution phenomenon occurs for multiple richer registered history packages.

Matched random-gene histories establish the finite specificity boundary. One of 16 matched banks reaches exact recovery by day 3, and 11 of 16 reach exact recovery by day 7. The frozen design contains 16 deterministic matched banks; these controls establish nonuniqueness of the history-recovery phenomenon rather than a high-resolution extreme-tail estimate. The exact factorization in Eq. (47) therefore establishes history sufficiency on the observed finite carrier together with a direct requirement for compression before a shared biological law can be assigned.

5 History compression and the boundary of pooled fate

Let H be an exact fate-sufficient history on a carrier L. Any quotient

$$q : \operatorname{im} H \to \mathcal{Q} \tag{48}$$

produces

$$Q = q \circ H. \tag{49}$$

The compression problem is

$$\min |\mathrm{im}\, Q| \quad \text{subject to} \quad \mathrm{Recovers}(Q, F_{14}). \tag{50}$$

The fate-defined oracle quotient gives the theoretical lower bound

$$|\mathrm{im}\, Q| \geq |\mathrm{im}\, F_{14}|\,. \tag{51}$$

The empirical question uses only fate-blind history coarsenings.

On the H037 carrier,

$$|L| = 748, \qquad |\mathrm{im}\, F_{14}| = 163. \tag{52}$$

Every exact registered fate-blind quotient remains injective. The first registered nontrivial compression is

$$748 \longrightarrow 747 \tag{53}$$

and creates the explicit mixed-fate collision

$$Q(\ell_1) = Q(\ell_2) = 22220 > 11112 > 02112, \tag{54}$$

with

$$F_{14}(\ell_1) = 2{:}1{:}0, \qquad F_{14}(\ell_2) = 0{:}0{:}1. \tag{55}$$

Hence

$$\boxed{\text{first registered pooled-history collision} \Longrightarrow \text{pooled terminal-fate ambiguity.}} \tag{56}$$

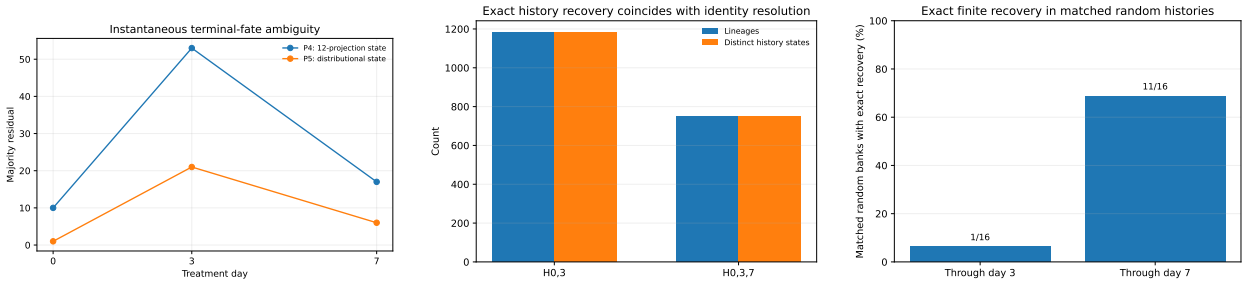
The H03 carrier yields the same qualitative boundary across a broader registered compression family. Exact pooled recovery remains tied to lineage-specific history resolution.

A restricted carrier of lineages with exactly concordant terminal replicate distributions contains a small noninjective exact quotient, establishing the feasibility of shared history classes when the terminal target is restricted to a replicate-stable subset. Its shared coverage is small and its matched permutation/random-history controls place it within common finite behavior. This observation motivates construction of the repeated-realization terminal object directly.

6 Repeated terminal realization and support-weighted fate relations

6.1 Paired terminal realizations

The Watermelon source defines two biological replicate experiments from the same starting Watermelon-transduced PC9 population. Each replicate underwent a separate 14-day osimertinib treatment, day-14 sorting into cycling, moderately cycling, and non-cycling populations, and single-cell RNA sequencing [23]. The experiments therefore provide separate treatment/culture realizations while retaining a common starting population and shared lineage-barcode ancestry. The replicate relation below measures agreement across these biological replicate realizations; it does not remove sampling variation or shared ancestry.



(a) Instantaneous Watermelon majority residual at days 0, 3, and 7 for two rich fate-blind state representations.

(b) Distinct accumulated-history states equal the number of lineages on the two exact-recovery history carriers.

(c) Frequency of exact terminal-fate recovery among 16 matched random-gene history banks.

Figure 3: Temporal state and accumulated history. Preterminal snapshots retain mixed terminal-fate fibers. Exact history factorization appears together with lineage-identifying history resolution and occurs frequently in matched high-dimensional random histories. Perfect finite prediction can therefore arise through lineage identification; the compression analysis tests whether that exactness survives representation shared across distinct lineages.

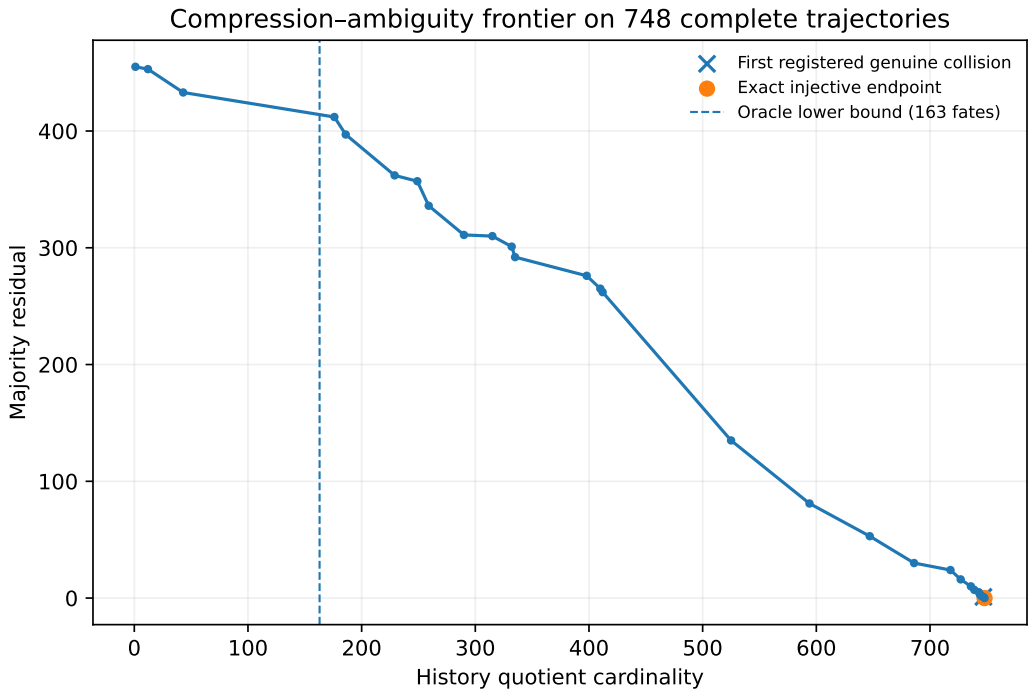


Figure 4: Compression–ambiguity frontier on the 748-lineage complete-trajectory carrier. The fate-defined oracle requires only 163 terminal classes, whereas every registered exact fate-blind history quotient remains injective. The first registered nontrivial compression, $748 \to 747$, produces majority residual 1 and the explicit mixed-fate collision reported in Eqs. (54)–(55).

Watermelon contains 1,550 lineages observed in both day-14 replicates. Their reduced terminal distributions satisfy

$$837 \text{ exactly concordant lineages}, \qquad 713 \text{ discordant lineages}. \tag{57}$$

The paired carrier contains

$$190 \tag{58}$$

distinct raw fate objects across the two replicates.

Define the direct symmetric relation

$$f \, R \, g \quad \Longleftrightarrow \quad \exists \ell : \{F_1(\ell), F_2(\ell)\} = \{f, g\}. \tag{59}$$

The graph has 265 distinct edges, including 254 nonself edges. Its equivalence closure produces

$$\boxed{190 \text{ raw fate objects} \longrightarrow 31 \text{ replicate-equivalence classes.}} \tag{60}$$

The largest class contains 128 raw fate objects. The lineage-pair discrimination fraction under this quotient is 0.0408754868.

The $k = 1$ closure has a canonical finite universal property: every terminal map that assigns equal values to the two registered replicate realizations of every paired lineage factors through this quotient. Biologically, this 31-class partition is the maximally permissive endpoint of the support-indexed filtration developed below. Of the 254 nonself direct edges, 201 have support one; the support-weighted hierarchy, rather than the $k = 1$ closure alone, is therefore the primary repeated-realization object.

6.2 History compression against the replicate-invariant target

Intersecting the paired terminal carrier with the complete H037 history carrier gives

$$540 \text{ lineages} \tag{61}$$

occupying 26 replicate-equivalence fate classes. A frozen fate-blind history quotient compresses this carrier as

$$540 \longrightarrow 511 \tag{62}$$

and preserves exact replicate-invariant fate:

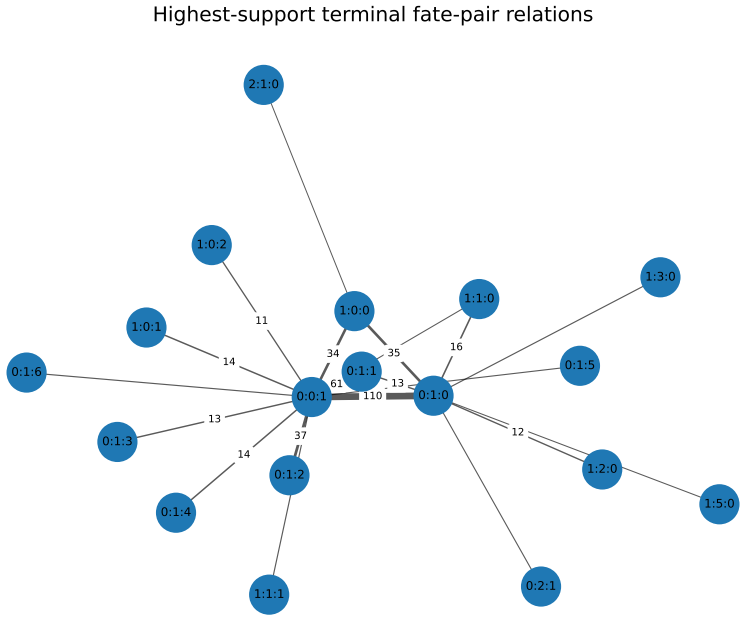
$$\text{Recovers}(Q_H, F_{\text{inv}}) = \text{true}. \tag{63}$$

Thirty-three lineages occur in non-singleton history fibers, and the largest shared history fiber has size 27. The same history quotient retains mixed fibers for replicate-1 fate, replicate-2 fate, and pooled terminal fate. Therefore

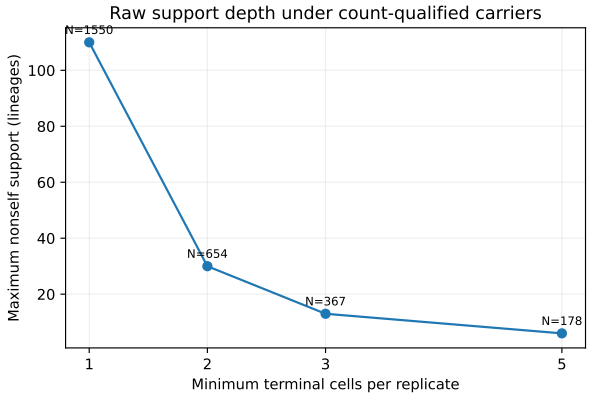
$$\boxed{Q_H \to F_{\text{inv}}} \tag{64}$$

exists as a nontrivial many-to-one finite factorization on this carrier, with realization-specific terminal detail retained at a finer level.

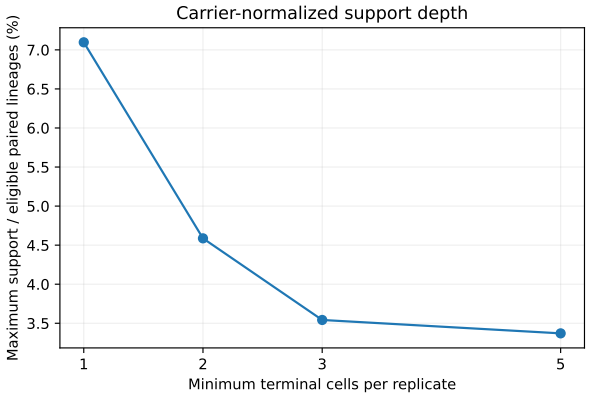
The exact factorization is common under matched finite controls: 180 of 1,000 fate-label permutations and 210 of 1,000 fiber-spectrum-matched random partitions are also exact; 7 of 16 matched random-gene histories recover the invariant target. The finite map in Eq. (64) therefore establishes a shared history class for the observed replicate quotient and motivates analysis of the terminal relation itself.



(a) Highest-support direct terminal fate-pair relations. Edge width scales with paired-lineage support; numeric labels mark the strongest relations.



(b) Raw maximum nonself support under minimum terminal-cell requirements. Carrier size contracts simultaneously.



(c) The same maximum support normalized by the number of eligible paired lineages: 7.10%, 4.59%, 3.54%, and 3.37%.

Figure 5: Repeated biological replicate terminal realization defines a weighted fate relation. The $k = 1$ 31-class closure is the maximally permissive endpoint of the persistence filtration; count-qualified sensitivity is reported in both raw support and carrier-normalized units.

7 Persistent replicate-support fate geometry

7.1 Weighted replicate graph

The direct replicate relation is weighted by the number of paired lineages supporting each raw-fate relation:

$$w(f,g) = \#\{\ell : \{F_1(\ell), F_2(\ell)\} = \{f,g\}\}. \tag{65}$$

Among the 254 nonself edges,

$$201 \tag{66}$$

have support one. The maximum nonself support is 110.

For each k, the threshold graph retains edges with $w \geq k$, and connected components define Π_k. The complete filtration extends from

$$|\Pi_1| = 31 \tag{67}$$

through increasingly fine partitions to the identity endpoint. Registered checkpoint levels include

k	1	2	3	5
$\lvert\Pi_k\rvert$	31	156	164	173
largest class	128	35	27	18.

(68)

Most of the size-128 $k = 1$ component separates at $k = 2$. A nested core persists through larger support values.

7.2 High-support core

The strongest direct relations are

$$\begin{aligned}
&0{:}0{:}1 \leftrightarrow 0{:}1{:}0, \quad w = 110,\\
&0{:}0{:}1 \leftrightarrow 0{:}1{:}1, \quad w = 61,\\
&0{:}0{:}1 \leftrightarrow 0{:}1{:}2, \quad w = 37,\\
&0{:}1{:}0 \leftrightarrow 1{:}0{:}0, \quad w = 35,\\
&0{:}0{:}1 \leftrightarrow 1{:}0{:}0, \quad w = 34.
\end{aligned} \tag{69}$$

The final nonself two-member core

$$\boxed{0{:}0{:}1 \leftrightarrow 0{:}1{:}0} \tag{70}$$

persists through support 110.

Lineage-level persistence follows the same filtration. Let

$$\sigma(\ell) = \max\{k : F_1(\ell) \sim_k F_2(\ell)\}. \tag{71}$$

The stable-lineage fraction decreases from one at $k = 1$ toward the raw-concordant fraction $837/1550 = 0.540$ at the identity endpoint.

7.3 Whole-profile null geometry

The persistence hierarchy is compared with 1,000 deterministic marginal-preserving replicate-pair permutations. Three prospectively frozen whole-profile summaries—class-resolution area, pair-discrimination area, and lineage-persistence area—each place the observed hierarchy in the registered upper tail

$$p = \frac{1}{1001}. \tag{72}$$

The direct edge-support spectrum has a different geometry: random pairings generate greater concentration in maximum direct support and repeated-edge count. The distinguishing empirical signal is therefore carried by the full connectivity/persistence profile; isolated edge multiplicities occupy the opposite tail of the permutation distribution.

7.4 Sampling-depth-conditioned replicate null

Terminal composition precision depends on the number of sequenced cells representing a lineage in each replicate. A revision-stage robustness null therefore preserves approximate paired sampling depth while randomizing replicate fate association. Replicate totals are assigned to the fate-blind bins

$$\{1\},\ \{2\},\ \{3,4\},\ \{5,\ldots,8\},\ \{9,\ldots,16\},\ \{17,\ldots,32\},\ \{33,\ldots\}, \tag{73}$$

and replicate-2 fate assignments are permuted only within the joint bin indexed by the observed pair (n_1, n_2). Across 1,000 deterministic permutations, the observed persistence profile remains in the upper tail. Corrected upper-tail probabilities are

$$p_{\text{class}} = \frac{4}{1001}, \qquad p_{\text{pair}} = \frac{19}{1001}, \qquad p_{\text{lineage}} = \frac{36}{1001}. \tag{74}$$

The persistence result therefore remains resolved after approximate conditioning on paired terminal lineage sampling depth (Fig. 6).

7.5 Alternative terminal-fate representation

The four-part composition representation described above yields 15 raw terminal states and maximum direct support 110. Its whole-profile values are

$$A_{\text{class}} = 81.000, \qquad A_{\text{pair}} = 94.1905, \qquad A_{\text{lineage}} = 86.9097. \tag{75}$$

All three exceed the corresponding 1,000 marginal-preserving permutation distributions with corrected upper-tail probability

$$p = \frac{1}{1001}. \tag{76}$$

Thus the support-indexed repeated-realization structure is retained under a substantially coarser composition coordinate whose equality relation is less sensitive to one-cell ratio changes.

7.6 History resolution across the persistent hierarchy

All frozen fate-blind history quotients are evaluated across the critical support levels. The history quotient that recovered the $k = 1$ replicate-invariant target remains exact on the stable-lineage subtype through $k = 62$ and fails at the identity endpoint $k = 111$. The explicit boundary witness contains two lineages with the same registered history

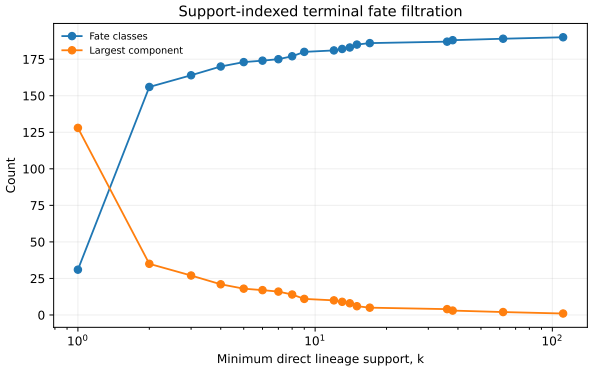
$$111111111111 > 111111111111 > 111111111111 \tag{77}$$

and terminal self-fates

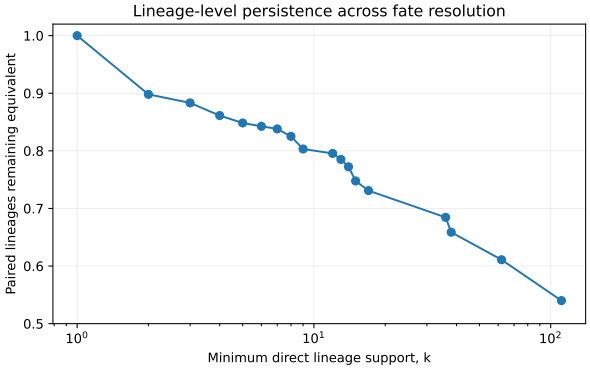
$$0{:}0{:}1 \qquad \text{and} \qquad 0{:}1{:}0. \tag{78}$$

On the full paired carrier, realization-class recovery by the same quotient is exact only at $k = 1$. Thirteen of 16 matched random history banks reach at least support 62 on the stable-subtype lane. Fine history recovery therefore remains a common property of matched high-dimensional temporal fingerprints.

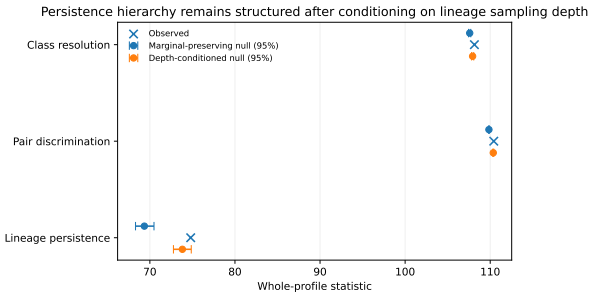
Persistent replicate-support fate geometry and robustness



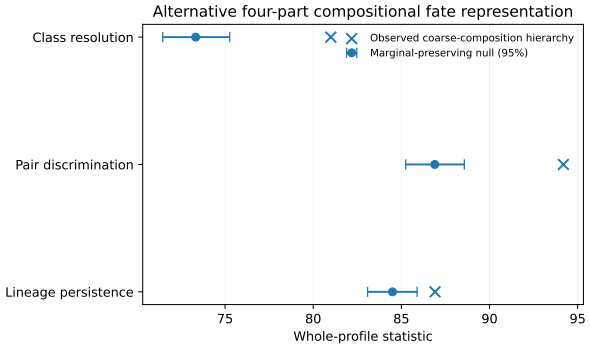
(a) Class count and largest connected component across the support-indexed terminal fate filtration.



(b) Fraction of paired lineages whose two terminal realizations remain equivalent at each support threshold.



(c) Observed whole-profile statistics against 95% intervals from the marginal-preserving and lineage-depth-conditioned nulls. Conditioned upper tails are 4/1001, 19/1001, and 36/1001.



(d) A coarse four-part composition endpoint retains the hierarchy relative to 1,000 marginal-preserving replicate-pair permutations; all three corrected upper-tail probabilities equal 1/1001.

Figure 6: The primary GCD-reduced hierarchy is strongly structured, remains resolved after approximate conditioning on paired lineage sampling depth, and is recapitulated under a substantially coarser terminal composition representation.

8 Terminal molecular inhabitation and localization

8.1 Molecular carrier and fate-blind state maps

The day-14 transcriptomic carrier contains 14,701 barcode-positive cells. Of these, 10,588 belong to the 1,550 paired terminal lineages, yielding 3,100 lineage–replicate molecular objects. Three fate-blind terminal molecular maps are constructed from a frozen 96-gene, 12-projection coordinate bank:

$$M^{\mathrm{abs}}, \qquad M^{\mathrm{cond}}, \qquad M^{\mathrm{cycle-red}}. \tag{79}$$

The first retains absolute terminal transcript state. The second fits molecular quantization within the cycling, moderate, and non-cycling sorting classes and omits the class label from the resulting

molecular state. The third reconstructs the fate-blind molecular state after exclusion of the registered cell-cycle coordinate bank.

Terminal molecular heterogeneity is extensive. Seventy-five of 190 raw terminal fates contain multiple lineage–replicate molecular objects. Exact same-lineage molecular objects are absent across terminal replicates in all three molecular lanes, and mean replicate distribution overlap is on the order of 10^{-5}–10^{-4}.

8.2 Molecular inhabitant of the high-support core

For the support-110 pair in Eq. (70), the exact common molecular supports are

$$\left|K^{\mathrm{abs}}\right| = 89, \qquad \left|K^{\mathrm{cond}}\right| = 190, \qquad \left|K^{\mathrm{cycle-red}}\right| = 170. \tag{80}$$

The corresponding lineage-weighted distributional overlaps are

$$0.021397, \qquad 0.050469, \qquad 0.040565. \tag{81}$$

Thus the high-support fate relation contains exact shared residual molecular states together with broad molecular heterogeneity.

At lower support, pairwise edge kernels remain nonempty. Whole-component intersections contract sharply. At support 61, endpoint-edge common-state counts are 17/35/34 across the three lanes, and only 1/6/1 states survive intersection over the full threshold-61 component. At support 37, endpoint-edge kernels contain 16/32/32 states and the full threshold-37 component has an empty common kernel in all three lanes.

8.3 Global molecular specificity controls

The complete persistence hierarchy shows higher within-component molecular profile area than matched-component and edge-support nulls. Fate reassignment and matched random gene banks give the controlling specificity result. For the conditioned and cycle-reduced whole-profile areas, matched random-gene-bank upper-tail values are 0.882353. The 16-bank design is interpreted as a finite matched-coordinate specificity screen rather than as a high-resolution tail estimate. The registered molecular control stack places the global persistence hierarchy within the matched molecular null distributions.

Lineage-level molecular reproducibility provides an independent boundary. Every paired lineage carries distinct replicate molecular objects in all three lanes. Spearman associations between lineage fate-persistence $\sigma(\ell)$ and replicate molecular overlap are small and negative:

$$-0.0610, \qquad -0.0863, \qquad -0.0673. \tag{82}$$

Persistent replicate-fate identity therefore remains distinct from exact transcriptomic realization at the measured terminal resolution.

8.4 Core localization

The support-110 common molecular states show substantial observed atlas exclusivity. Raw-fate-exclusive counts are

$$74/89, \qquad 140/190, \qquad 124/170 \tag{83}$$

for the absolute, conditioned, and cycle-reduced lanes. All exclusive states occur in more than one lineage; replicate-balanced exclusive counts are 40, 70, and 69. The union of these exclusive states covers 6.58%, 12.04%, and 10.52% of support-110 component lineage–replicate objects, respectively.

The specificity controls assign the interpretation. Under fate reassignment, the support-110 exclusive-count upper tail is 1.0 in every molecular lane. Matched random-bank upper tails are 0.6471, 0.1765, and 0.2353. Component/background localization contrast likewise lacks a distinctive upper-tail position under the complete control stack. The support-110 kernel is therefore a real observed common molecular repertoire whose atlas localization is compatible with the finite occupancy structure generated by reassigned fates and matched molecular banks.

The nested root components sharpen from three fates at support 38 to two fates at support 62, with the support-62 and support-110 components equal. Their common molecular kernels expand under component refinement, as required by set-intersection monotonicity. Molecular specificity decreases with this refinement: exclusive fractions decline from 5/6 to 140/190 in the conditioned lane and from 1 to 124/170 in the cycle-reduced lane, accompanied by increased outside-fate prevalence. The persistent fate hierarchy therefore resolves toward broader molecular sharing at the terminal transcriptomic level.

8.5 Independent molecular and chromatin inhabitation

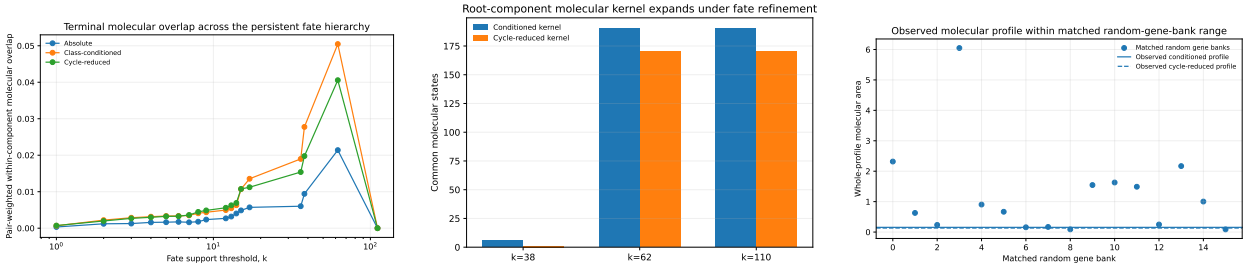
The independent hematopoietic carrier provides a second terminal realization test with a fourteen-label source-native fate alphabet and orthogonal RNA and chromatin modalities [34]. Among 1,453 clones linking day-2.5 RNA state to day-5 fate, 882 realize more than one pooled terminal fate. Across the non-singleton terminal relation components, the terminal transcriptomic analysis yields nonempty molecular kernels in

$$172 \text{ of } 396 \tag{84}$$

component–representation objects, with maximum kernel size 108. The registered fate-reassignment, matched-component, matched-random-bank, and state-prevalence controls do not establish a distinctive component-specific molecular identity.

Chromatin accessibility supplies an orthogonal terminal inhabitant. RNA and ATAC share 1,689 terminal clones, and 84 RNA-defined non-singleton relation components possess nonempty chromatin kernels. Because the RNA and ATAC measurements arise from different destructively sampled cells within the same clone, this result establishes clone-level cross-modality inhabitation rather than same-cell multiomic identity. Together with the Watermelon analysis, the terminal result is therefore stable at the level of shared molecular or chromatin inhabitants while remaining heterogeneous under the corresponding specificity controls.

Terminal molecular realization of persistent fate geometry



(a) Within-component terminal molecular overlap across fate-support levels for absolute, class-conditioned, and cycle-reduced molecular representations.

(b) Common molecular-state counts in the nested high-support root component. Component refinement expands the exact common repertoire.

(c) Whole-profile molecular area for 16 matched random gene banks, with the observed conditioned and cycle-reduced values shown as references.

Figure 7: The high-support core contains exact residual shared states, while whole-profile and core-localization specificity remain within the matched molecular-control range.

9 Independent tests of multimodal state and temporal history

The preceding analyses identify three recurring possibilities: a state representation can remain fate-ambiguous, become exact by resolving lineage identity, or retain a many-to-one state partition that is exactly fate-sufficient. The CellTag-multi hematopoietic and reprogramming systems provide independent stress tests of these possibilities using different biological processes, terminal fate alphabets, orthogonal modalities, and temporal designs [34].

9.1 Orthogonal state completion in hematopoiesis

The reconstructed hematopoietic authority contains 1,453 clones with admitted day-2.5 RNA state and day-5 future fate. Its pooled terminal fate is a distribution over fourteen source-native hematopoietic labels. The richest registered RNA state contains 1,065 state fibers, of which 168 remain mixed for pooled fine fate and expose 2,564 discordant clone pairs. Richer transcriptomic state therefore contracts ambiguity substantially while preserving explicit future multiplicity.

A strict multimodal carrier contains 201 clones with day-2.5 RNA and ATAC state together with future fate. At the deepest registered state level, RNA alone occupies 184 state classes and leaves 32 fine-fate-discordant clone pairs; ATAC occupies 195 classes and leaves 11. Their exact Cartesian product occupies

$$201 \text{ states for } 201 \text{ clones} \tag{85}$$

and removes all observed fine-fate collisions. Exact fine-fate recovery therefore occurs at clone identity on this carrier. The immediately shallower joint representations retain a fine-fate collision, so the observed endpoint supplies a direct multimodal example of exactness achieved through individuality rather than a many-to-one fine-fate law.

Noninjective exact surfaces do exist for coarser futures. For example, the majority-fate joint state reaches 199 states for 201 clones, and an ATAC state reaches 86 states for 89 clones on the arm-invariant target. Their matched sham, fiber-spectrum, random-bank, representation, and count controls place these surfaces inside common finite behavior. The independent hematopoietic result therefore localizes the contribution of multimodal completion: orthogonal state resolves additional clone individuality and future ambiguity, while control-supported shared exact fine-fate structure remains absent.

9.2 Temporal history across biological replicate reprogramming

The reprogramming authority contains 42,081 clone-table cells and 8,050 reconstructed clones across two biological replicates. Day-21 future fate uses the source-native categories *dead-end*, *reprogramming*, and *transition*. On the larger RNA history carriers, the deepest day-3 snapshot leaves fine-fate residuals 18 and 10 in replicates 1 and 2; incorporating day-12 history reduces these to 5 and 0. Replicate-2 RNA exactness coincides with a 159-state partition on 159 clones.

ATAC histories recover fine fate exactly in both biological replicates, with 27 states on 27 clones and 35 states on 35 clones. The strict RNA–ATAC history carrier is smaller still—17 and 24 clones—and its day-3 multimodal snapshot is already clone-injective and exact in both replicates. Ordered history therefore adds no result-grade information on that strict carrier. Temporal-sham, matched-random-history, and matched-fiber controls each place the observed strict-multimodal gain at the corresponding null boundary rather than establishing temporal specificity.

The biological replicates also provide a transport audit. At package depth 5, the RNA ordered-history representation contains two state values shared between biological replicates; neither shared value preserves fine day-21 fate. The corresponding ATAC and strict multimodal ordered-history representations contain no shared depth-5 state values across replicates. The RNA comparison therefore supplies a finite nontransport witness on the observed shared support, whereas the ATAC and multimodal comparisons are authority-limited by the absence of recurring state values. Within-replicate exactness consequently does not by itself establish transport of a state–fate rule between biological realizations.

10 Cross-carrier fate sufficiency at the identity boundary

The independent carriers motivate a common coordinate system for asking whether future resolution exceeds the individuality introduced by state refinement. The synthesis freezes every eligible registered state/future surface from the preceding analyses and adds no new carrier, feature, clustering, fate quotient, or outcome-tuned search. The resulting authority contains 958 eligible surfaces distributed across 54 carrier/future families; 849 distinct state partitions remain after within-family deduplication, with 27 typed exclusions outside the finite state/future-map comparison.

Table 1: Independent stress tests of the finite relational calculus. "Exact" refers to carrier-scoped factorization under the registered state and future representations.

System	State authority	Future authority	Principal geometry
Watermelon PC9	RNA snapshots, multitime history	day-14 distributional fate; biological replicate realization	Snapshot mixed fibers; history exactness at lineage-identifying resolution; nonrandom support filtration; heterogeneous terminal molecular repertoire.
CellTag LSK	day-2.5 RNA and ATAC, same clone across destructive measurements	day-5 fourteen-label fate; two technical fate arms	RNA and ATAC retain fine-fate collisions separately; RNA×ATAC removes them at 201/201 clone identity; noninjective exact coarse surfaces are not control-specific.
CellTag iEP	day-3 and day-12 RNA/ATAC histories in two biological replicates	day-21 three-category fate	RNA history improves ambiguity; ATAC and strict multimodal exactness are fingerprint-level. Two shared depth-5 RNA history values fail fine-fate transport; ATAC and multimodal depth-5 transport are authority-limited by zero shared state values.

10.1 Identity resolution and future-discordance resolution

For a carrier of N lineages, let

$$T = \binom{N}{2} \tag{86}$$

be the number of unordered lineage pairs. For a state representation Q, let $C(Q)$ be the number of distinct lineage pairs that collide in state,

$$C(Q) = \#\{\{x, y\} : x \neq y,\ Q(x) = Q(y)\}, \tag{87}$$

and let $D(Q, F)$ be the number of those state-colliding pairs whose registered futures disagree,

$$D(Q, F) = \#\{\{x, y\} : x \neq y,\ Q(x) = Q(y),\ F(x) \neq F(y)\}. \tag{88}$$

Let $D_0(F)$ denote the total number of fate-discordant lineage pairs on the carrier. Identity resolution and future-discordance resolution are

$$I(Q) = \frac{T - C(Q)}{T}, \qquad J(Q, F) = \frac{D_0(F) - D(Q, F)}{D_0(F)}. \tag{89}$$

The fate-selective surplus is

$$\boxed{S(Q, F) = J(Q, F) - I(Q).} \tag{90}$$

Thus $S > 0$ records a representation that resolves fate-discordant pairs more strongly than lineage pairs in general. Under a fate-independent partition of a fixed carrier, reductions in fate-discordant collisions are expected to track generic lineage-pair resolution; the familywise controls therefore determine whether a positive surplus exceeds the geometry produced by partition refinement

alone. On this scale, $S = 0.057$ means approximately 5.7 percentage points more resolution of fate-discordant lineage pairs than generic lineage-pair resolution supplied by the same representation. When $C(Q) > 0$, the remaining conditional future ambiguity is

$$A(Q, F) = \frac{D(Q, F)}{C(Q)}. \tag{91}$$

These quantities preserve the finite semantics of the earlier fiber tests while placing snapshot, history, multimodal, and persistent-state representations in one cross-carrier geometry.

10.2 Exact recovery concentrates at lineage identity

Exact recovery is equivalent to

$$D(Q, F) = 0, \tag{92}$$

whereas injectivity is equivalent to

$$C(Q) = 0. \tag{93}$$

The canonical exact-recovery surfaces separate accordingly into

$$\boxed{\text{352 injective exact surfaces,} \qquad \text{18 noninjective exact surfaces.}} \tag{94}$$

The 18 noninjective exact surfaces are the complete registered candidates for a shared exact fate law in the synthesized bank. After applying the corresponding familywise fate-permutation and matched-fiber controls,

$$\boxed{\text{18 shared-exact candidates} \longrightarrow \text{0 control-qualified shared-exact laws.}} \tag{95}$$

Exact future recovery therefore concentrates at the lineage-identity boundary across the registered representation class. This boundary remains finite rather than asymptotic: near-complete identity can still retain $C > 0$ and $D > 0$, so $I(Q) \approx 1$ does not itself imply exactness.

10.3 Future alignment below exactness

The identity boundary coexists with a positive non-exact result. Fate-selective surplus reaches

$$S = 0.0568107 \tag{96}$$

in the positive-resistant precursor family, $S = 0.0212772$ in the hematopoietic arm-invariant family, and $S = 0.0231970$ in the first reprogramming biological replicate. Familywise control qualification is retained in the Watermelon/Rewind program and the CellTag reprogramming system. The hematopoietic maximum is retained descriptively without the same control qualification.

The cross-carrier synthesis therefore distinguishes two empirical regimes:

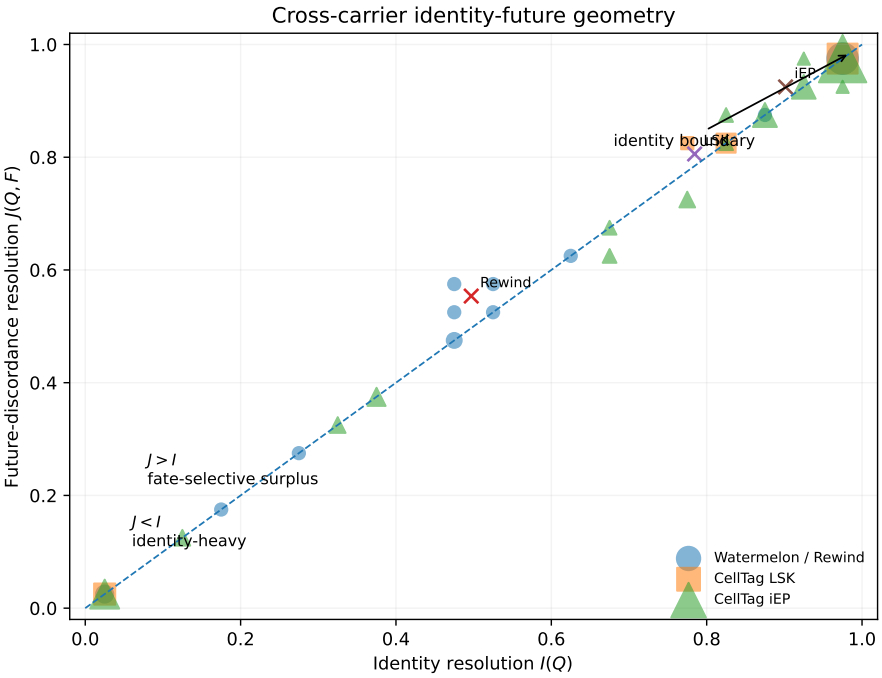
$$\boxed{\begin{array}{c} D = 0 \text{ occurs predominantly at lineage identity,} \\ S > 0 \text{ can occur with } D > 0. \end{array}} \tag{97}$$

Future alignment can therefore be measured below the exactness boundary. This is the finite form of constrained future realization: a state representation can preferentially remove future-discordant collisions while preserving residual ambiguity among lineages that remain state-indistinguishable.

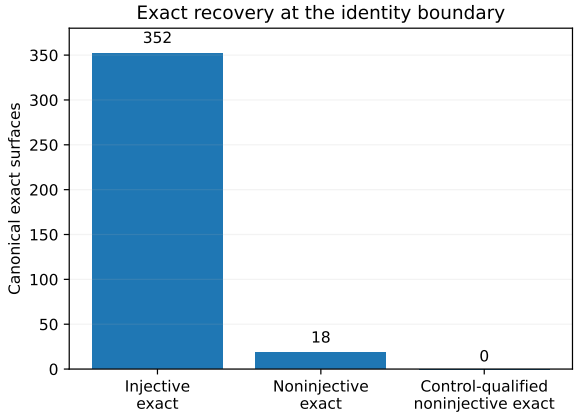
10.4 Programmatic closure of the registered representation class

The synthesis also supplies a stopping criterion. The independent hematopoietic carrier, orthogonal chromatin state, multimodal completion, temporal history, biological replicate history, persistent terminal geometry, and terminal molecular analyses all remain expressible in the existing relational calculus. The registered exact noninjective candidates do not yield a control-qualified shared exact law, whereas control-qualified future-selective surplus survives in more than one system group. Under the frozen plateau criteria, the registered representation program therefore reaches an empirical boundary: further progress within this program is more likely to require a new carrier, measurement class, intervention, or experimental authority than continued refinement of the registered representation family.

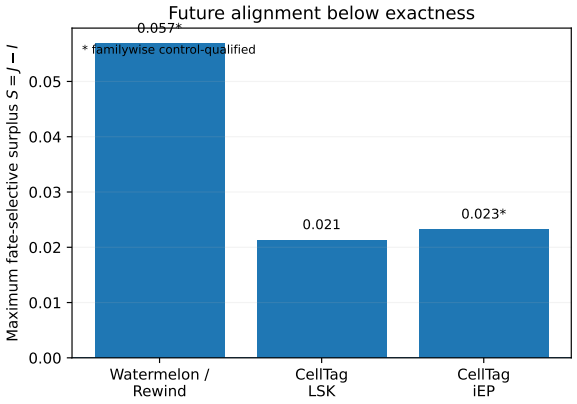
Identity–future geometry across the registered biological program



(a) Binned density of canonical decision-eligible state/future surfaces in the identity–future plane. Marker area is proportional to the number of surfaces in each bin; crosses identify the maximum fate-selective-surplus surface in each system group. The diagonal is $J = I$.



(b) Canonical exact-recovery census. Exactness is dominated by injective lineage fingerprints; none of the 18 noninjective exact candidates survives the combined familywise controls.



(c) Maximum fate-selective surplus $S = J - I$ by synthesis system group. Asterisks mark groups with familywise control-qualified positive surplus.

Figure 8: Cross-carrier fate sufficiency at the identity boundary. Exact recovery accumulates at lineage-identifying representations, while positive fate-selective surplus remains measurable below exactness and is control-qualified in two system groups.

11 Unified finite relational calculus

The complete calculus is organized by a sequence of finite questions. A measured state map $P_t : \mathcal{L}_t \to \mathcal{S}_t$ defines state fibers and their future images. Temporal accumulation produces $H_{\leq t}$; fate-blind quotienting tests whether exactness survives a many-to-one representation; repeated terminal realization generates a support-weighted relation and the persistence filtration $\{\Pi_k\}$; molecular support assigns each terminal fate a finite repertoire $\mathcal{M}(f)$; and cross-carrier synthesis compares the resolution of lineage identity with the resolution of future discordance.

The resulting inferential chain can be summarized as

$$\boxed{\begin{array}{c}\text{state fiber geometry} \longrightarrow \text{history sufficiency} \longrightarrow \text{shared compression} \\ \longrightarrow \text{repeated-realization persistence} \longrightarrow \text{molecular inhabitation} \\ \longrightarrow \big(I(Q), J(Q,F), S(Q,F)\big).\end{array}} \tag{98}$$

Each stage preserves the carrier and authority needed to interpret the next. Exact recovery determines whether the registered future is a function of the measured representation. Injectivity determines whether that exactness is supplied by lineage identity. Compression asks whether multiple lineages share an exact state–future representation. Repeated realization asks which terminal distinctions survive independent or parallel realization. Molecular support asks how those persistent terminal relations are inhabited. The identity–future geometry finally asks whether a representation resolves future discordance more strongly than its generic lineage-identity resolution.

The cross-carrier result supplies the calculus with a terminal empirical boundary:

$$\boxed{\begin{array}{c}\text{exact recovery concentrates at lineage identity,} \\ \text{control-qualified future alignment exists below exactness,} \\ \text{persistent fate can admit multiple molecular realizations.}\end{array}} \tag{99}$$

The phrase *constrained future realization* denotes this combined geometry. It refers to preferential contraction of future ambiguity under measured state, history, modality, and repeated realization while preserving the experimentally observed distinction between future alignment, exact fate sufficiency, lineage identity, and molecular sameness.

12 Discussion

This study develops and applies a finite relational calculus for lineage-resolved state–fate inference across cancer resistance, hematopoietic differentiation, and cellular reprogramming. The central objects—state fibers, temporal histories, noninjective quotients, repeated terminal relations, persistence filtrations, and molecular-support intersections—separate biological claims that are frequently compressed into a single language of prediction. The cross-carrier synthesis extends that separation by placing every eligible representation in a common identity–future geometry.

The first recurring result is that measured state can be strongly future-informative while retaining explicit future multiplicity. ReSisTrace and Rewind show this in pre-treatment cancer state; Watermelon shows it across multiple preterminal times; the hematopoietic and reprogramming

carriers show it again under different fate alphabets and state modalities. Exact finite recovery is therefore a stronger statement than enrichment, reduced residual, or high classification performance. A mixed state fiber remains a direct witness that the registered future is not a function of that state representation on the observed carrier.

The second result concerns exactness itself. Watermelon history becomes exact when the registered history approaches lineage identity. RNA–ATAC completion in the hematopoietic carrier removes all observed fine-fate collisions only when the joint state becomes 201-for-201 clone-injective. In the reprogramming system, ATAC histories are exact at 27/27 and 35/35 clone identity, while the strict multimodal snapshot is already exact and injective before temporal accumulation. The cross-carrier census places these examples in a common boundary: 352 canonical exact surfaces are injective, 18 are noninjective, and none of the 18 noninjective exact candidates survives the combined familywise controls. Exact prediction therefore frequently records identity resolution rather than a compact state–fate law shared across lineages.

The compression analyses explain why this distinction matters. Exact recovery establishes sufficiency on a finite carrier. A noninjective fate-blind quotient asks whether that sufficiency persists when distinct lineages are required to share one representation. The Watermelon compression frontier makes the distinction explicit: the first registered nontrivial pooled-history collision creates terminal-fate ambiguity. The independent CellTag carriers add complementary evidence: some noninjective exact coarse surfaces exist, yet the corresponding sham, matched-fiber, representation, and random-bank controls do not establish a robust shared exact law. The cross-carrier control census therefore promotes compression from a local diagnostic to a general inferential boundary.

Repeated realization contributes a different form of biological order. In Watermelon, the terminal relation generates a support-indexed hierarchy whose whole-profile geometry is strongly separated from marginal-preserving nulls and remains resolved after conditioning on lineage sampling depth and after replacing the primary GCD-reduced fate coordinate with a coarser composition representation. CellTag LSK shows related same-clone terminal variation across technical fate arms, while its topology remains authority-limited for biological replication. The iEP system supplies two biological replicates and separates transport failure from transport authority. At depth 5, two RNA ordered-history state values recur across replicates and neither preserves fine fate, while the ATAC and strict multimodal histories have no shared state values at that depth and therefore do not support a direct transport test. Reproducibility has to be evaluated at the level of the recurring state relation or hierarchy being claimed rather than inferred from within-carrier exactness.

Terminal molecular structure supplies a final separation. Watermelon persistent fate components contain exact shared molecular states and a high-support residual kernel, while whole-profile and localization controls place those states within matched molecular null geometries. The independent hematopoietic carrier similarly contains transcriptomic kernels in 172 of 396 component–representation objects and chromatin kernels in 84 RNA-defined non-singleton components, without control-supported unique molecular localization. A reproducible fate relation can therefore be inhabited by multiple terminal molecular configurations, including orthogonal transcriptomic and chromatin realizations at clone level.

The cross-carrier identity–future synthesis provides the program's strongest positive result below exactness. Fate-selective surplus $S = J - I$ measures future-discordance resolution after subtracting the amount of generic lineage-identity resolution supplied by the representation. Positive surplus is familywise control-qualified in the Watermelon/Rewind program and in CellTag iEP. Measured state can therefore resolve future discordance more strongly than generic lineage-identity resolution

while residual ambiguity remains. This result gives a finite empirical meaning to constrained future realization: biological information can preferentially narrow accessible futures without yielding a universal exact state–fate law.

The same synthesis defines a plateau under the frozen criteria for the registered representation class tested here. The program spans transcriptomic snapshot state, temporal history, chromatin accessibility, multimodal state, technical parallel realization, biological replicate realization, persistent fate geometry, terminal molecular support, and cross-modality inhabitation. Every discovered structure remains representable by the existing relational calculus, while the registered search bank yields no control-qualified shared exact fate law below the identity boundary. Under these criteria, further progress within this program is more likely to require a genuinely new carrier, measurement class, intervention, or experimental authority than continued refinement of the registered representation family.

These distinctions have direct methodological implications for translational state–fate studies. A biomarker that enriches for an outcome, a representation that becomes exact by identifying individual lineages, a compressed representation that remains exact across multiple lineages, and a state coordinate with positive fate-selective surplus support different biological claims. Resistance-screening and biomarker-development studies can therefore use the calculus to distinguish predictive association, carrier-scoped sufficiency, shared fate structure, replicate persistence, and terminal molecular realization without assigning all predictive success the same mechanistic meaning.

The contribution is consequently both empirical and formal. Empirically, the study identifies an identity boundary for exact future recovery together with control-qualified future alignment below that boundary. Formally, it supplies a finite language for determining what kind of state–fate knowledge an experiment has actually acquired. The field commonly asks how well state predicts fate; the calculus developed here additionally asks whether that prediction reflects individuality, a shared representation, persistent realization structure, or a molecularly distinctive terminal state.

13 Methods

13.1 Data sources and authority

The analysis uses published public data and source materials from cancer resistance, hematopoietic differentiation, and cellular reprogramming experiments. The FGF2/imatinib CML system is taken from Traer *et al.* [24] and is used for the measurement-authority analysis. ReSisTrace uses the public ovarian-cancer sister-cell lineage experiment of Dai *et al.* [22]. Rewind uses the melanoma Carbon Copy/barcode system of Emert *et al.* [21], together with its source relation to the rare-state resistance program characterized by Shaffer *et al.* [20]. Watermelon uses the public PC9 osimertinib time course of Oren *et al.* [23]. The independent hematopoietic and fibroblast-to-endoderm-progenitor reprogramming carriers use the public CellTag-multi experiment of Jindal *et al.* [34], including its public scRNA-seq, scATAC-seq, clone authority, and source code.

Each analytical carrier is reconstructed from source-owned cell identities, barcode or sister relations, sample time, perturbation, replicate, and terminal outcome coordinates. Missing lineage-time observations remain missing. Lineage relations are ancestry relations among sampled descendants and are kept separate from same-cell longitudinal identity.

13.2 Finite lineage carriers

A lineage carrier is a finite set $\mathcal{L}$. When barcode identity is multi-valued, the complete registered barcode set defines the canonical lineage identity. At time t, the observed carrier is

$$\mathcal{L}_t = \{\ell : \ell \text{ has a registered observation at } t\}. \tag{100}$$

History analyses use fixed-support intersections such as

$$\mathcal{L}_{0,3,14} = \mathcal{L}_0 \cap \mathcal{L}_3 \cap \mathcal{L}_{14} \tag{101}$$

and

$$\mathcal{L}_{0,3,7,14} = \mathcal{L}_0 \cap \mathcal{L}_3 \cap \mathcal{L}_7 \cap \mathcal{L}_{14}. \tag{102}$$

These fixed-support carriers prevent missingness pattern from functioning as an implicit fate coordinate.

13.3 State maps and fate-blind coordinate banks

For this manuscript, a *registered* analysis object is a repository-tracked specification whose content hash and commit precede the corresponding decision computation. Tranche-level plans for state packages, control families, quotient searches, and null metrics were frozen within the analysis workflow before the associated outcome decision was evaluated. These records document fate-blind construction and retrospective boundaries; they are not presented as an external public preregistration of the complete project.

Every decision-eligible state map is constructed independently of terminal fate labels. Transcriptomic states use deterministic normalization, frozen gene banks, deterministic projection grouping, and finite quantization. Source-highlighted retrospective signatures are retained as interpretive coordinates and excluded from independent state-to-fate decision surfaces when their selection used the terminal outcome.

For the Watermelon temporal carrier, the same state-map architecture is applied at days 0, 3, and 7. Lineages containing multiple cells at one time retain a finite within-lineage state distribution or set-valued state representation according to the registered package. Decision-eligible molecular state values exclude barcode identity, sample replicate, terminal cell count, and missingness.

13.4 Recovery, residual, and explicit witnesses

Recovery is evaluated by Eqs. (10)–(13). Every failed exact-recovery decision is accompanied by at least one explicit pair (ℓ_1, ℓ_2) satisfying Eq. (5). Majority residual is used as a finite ambiguity measure. Fate permutations preserve the appropriate carrier structure and fate multiplicities.

13.5 History compression

History candidates are deterministic fate-blind functions of the registered state trajectory. Compression families include coordinate deletion, quantization coarsening, transition signatures,

route motifs, heterogeneity summaries, temporal algebra, and time-order controls. Exact recovery and quotient cardinality are recorded jointly. An exact quotient is classified as genuinely compressed only when

$$|\mathrm{im}\, Q| < |\mathcal{L}|\,. \tag{103}$$

The fate-defined oracle quotient is retained only as a theoretical lower bound.

13.6 Replicate equivalence

For the paired Watermelon terminal carrier, the direct replicate relation and edge support are defined by Eqs. (59) and (21). The canonical replicate quotient is the equivalence closure of the direct relation. Support-indexed relations use Eq. (22); connected components are recomputed at every observed critical support value. Pairwise bottleneck persistence uses Eq. (24). Marginal-preserving replicate-pair permutations keep the replicate-1 and replicate-2 fate multisets fixed and randomize their cross-replicate pairing.

The three whole-profile summaries integrate complementary aspects of this filtration across support thresholds. The *class-resolution area* summarizes how strongly the raw fate space resolves into distinct equivalence classes as support requirements increase. The *pair-discrimination area* summarizes separation of raw-fate pairs across the filtration. The *lineage-persistence area* summarizes how deeply paired lineages retain equivalence between their two terminal realizations. Larger values therefore correspond, respectively, to greater fate-space resolution, greater pairwise discrimination, and replicate-linked lineage agreement persisting to stronger support thresholds. Exact discrete definitions and normalization conventions are included in Supplementary Note 7.

Two revision-stage robustness analyses were specified after internal manuscript review and are reported separately from the frozen primary analysis. The sampling-depth-conditioned null uses the terminal-count bins in Eq. (73) and permutes replicate-2 fate assignments within each joint (n_1, n_2) bin for 1,000 deterministic iterations. The alternative terminal representation maps each three-class count vector to a four-part integer simplex composition by largest-remainder apportionment and recomputes the complete support filtration and 1,000 marginal-preserving replicate-pair permutations. These analyses test sampling-depth and endpoint-discretization sensitivity without changing the primary GCD-reduced fate definition.

13.7 CellTag multimodal and temporal carriers

For hematopoiesis, day-2.5 RNA and ATAC state are reconstructed at clone level from distinct destructively sampled cells. No synthetic same-cell pairing is introduced. RNA and ATAC state packages are finite, fate-blind representations fitted entirely from the preterminal time point. Their Cartesian product is the joint state. Day-5 future fate retains the complete source-native fourteen-label distribution, with source-grouped, majority, breadth, arm-specific, and arm-invariant futures analyzed as registered coarsenings. The two day-5 arms are technical fate arms and are not interpreted as biological replicates.

For reprogramming, day-3 and day-12 RNA and ATAC observations define snapshot, ordered-history, unordered-history, reversed-history, and transition-signature representations independently within each biological replicate. The day-21 future alphabet is the source-native three-category fate. Strict multimodal carriers require clone-linked RNA and ATAC observations at the registered preterminal

times and a terminal future; same clone denotes lineage continuity across destructive samples rather than same-cell longitudinal measurement.

13.8 Cross-carrier identity–future synthesis

Every synthesis-eligible finite state/future map is reconstructed from the frozen predecessor artifacts. Within each carrier/future family, duplicate state partitions are removed by exact partition hash while preserving a typed exclusion registry for authority boundaries and non-map summary objects. For N lineages, $T = \binom{N}{2}$ is the unordered-pair count; C counts state-colliding lineage pairs; D counts fate-discordant state collisions; and D_0 counts all fate-discordant lineage pairs. Identity resolution, future-discordance resolution, fate-selective surplus, and conditional future ambiguity use Eqs. (89)–(91).

Familywise fate-label controls preserve the registered carrier and future multiplicities while randomizing future assignment. Matched-fiber controls preserve the state-partition geometry relevant to the candidate exact surface. A noninjective exact surface is classified as a control-qualified shared exact law only when it passes both registered control families. Positive fate-selective surplus is qualified at the family level using the prospectively frozen family maximum rather than treating each surface or refinement edge as an independent test. Cross-replicate state–fate transport is evaluated only when state values have matching registered semantics in both biological replicates. The audit records both the number of state values shared across replicates and the number of those shared values that preserve the registered future, so absence of recurring state values remains an authority limitation rather than evidence of transport failure.

13.9 Terminal molecular states

The primary terminal molecular bank contains 96 fate-blind genes divided deterministically into twelve eight-gene projections. Absolute states are quantized on a common terminal reference. Class-conditioned states use thresholds fitted separately within cycling, moderate, and non-cycling cells and omit the sorting-class identity from the resulting state. Cycle-reduced states exclude the registered S/G2M/cell-cycle coordinate bank and reconstruct a matched 96-gene fate-blind bank.

Molecular support is computed first at the lineage–replicate level and then aggregated to raw fates with equal lineage–replicate weighting. For a persistent fate component C, common support and exclusive support use Eqs. (26) and (27). Matched-component permutations, fate reassignments, edge-support permutations, matched random-gene banks, replicate swaps, lineage-size sensitivity surfaces, and state-prevalence controls define the molecular specificity stack.

13.10 Formal verification and computational reproducibility

The finite definitions used in the paper—recovery as fiber constancy, explicit nonrecovery witnesses, fate-sufficient compression, replicate-equivalence closure, nested quotient maps, support-threshold filtration, stable-lineage subtypes, molecular-support intersection, identity resolution, fate-discordant collision resolution, fate-selective surplus, and evidence boundaries—were implemented and machine checked in Lean 4 within the private research environment used to develop the analysis. Empirical carrier constants, source hashes, deterministic analysis artifacts, and control outputs are

audited independently by executable validation scripts. The formal verification and the public empirical-reproduction surface are therefore distinct: the former checks the finite propositions used by the analysis, while the latter allows readers to reproduce the numerical results reported in this manuscript. The private formal-development environment is not required for public empirical reproduction and is not distributed with this paper.

14 Data and code availability

The source datasets analyzed here are publicly available from the repositories associated with the original studies, including GEO accessions reported by ReSisTrace, Rewind, Watermelon, and CellTag-multi [22, 21, 23, 34]. Raw third-party source data are not mirrored in the manuscript repository unless their redistribution terms explicitly permit it. The public release instead records source accessions, source-code authority, acquisition metadata, hashes where applicable, and deterministic acquisition instructions.

Paper-specific reproducibility materials are available at

https://github.com/zedjames/relational-cell-fate

The repository contains the frozen empirical result tables, derived manuscript-facing summaries, public-source provenance, paper-specific numerical validation code, control outputs, and figure-regeneration scripts needed to reproduce and inspect the results reported here. It intentionally excludes the private development repository, its broader formal architecture, generic research-generation machinery, and reusable implementation of the relational calculus.

The archival record assigned to this manuscript has DOI 10.5281/zenodo.22969066. The DOI-backed release is the immutable computational companion to this preprint; later development of the GitHub repository does not alter the archived release associated with this manuscript.

15 Scope of the present study

The manuscript applies the finite relational calculus to public lineage-resolved carriers spanning transcriptomic state, chromatin accessibility, temporal history, multimodal state, terminal distributional fate, technical parallel realization, biological replicate realization, persistent fate support, and terminal molecular or chromatin inhabitation. The principal statements concern exact factorization, identity resolution, finite ambiguity, quotient structure, fate-selective surplus, repeated realization, cross-replicate transport, and molecular localization. Under the frozen plateau criteria, the cross-carrier synthesis establishes an empirical boundary for the registered representation class; causal mechanisms, intervention-specific transition laws, transfer to primary human tumors, untested transformations of the existing measurements, and any universal biological state–fate law remain outside the present authority.

Supplementary Information and reproducibility materials

Complete supplementary analyses, extended result tables, source and carrier provenance, control outputs, cross-carrier synthesis artifacts, validation receipts, and figure-generation materials are maintained in the public reproducibility repository accompanying this work:

https://github.com/zedjames/relational-cell-fate

The DOI-backed archival record corresponding to this manuscript is identified by:

https://doi.org/10.5281/zenodo.22969066

Author contributions

Zed James conceived the study, developed the finite relational calculus and analysis framework, reconstructed the public lineage-resolved carriers, designed and performed the computational analyses and control studies, carried out the formal verification, interpreted the results, and wrote the manuscript.

AI assistance disclosure

OpenAI ChatGPT was used as a computational and editorial assistant for manuscript formatting and language-level editing, LaTeX preparation, figure preparation, code and script drafting, execution and troubleshooting of analysis and reproducibility scripts, and organization of the public reproducibility materials. The author defined the scientific questions and analysis criteria, reviewed the underlying data and computational outputs, interpreted the results, and assumes full responsibility for the manuscript and its conclusions.

Competing interests

The author declares no competing interests.

References

[1] C. H. Waddington, *The Strategy of the Genes*, George Allen & Unwin, London (1957).

[2] L. Kester and A. van Oudenaarden, "Single-Cell Transcriptomics Meets Lineage Tracing," *Cell Stem Cell* **23**, 166–179 (2018), doi:10.1016/j.stem.2018.04.014.

[3] C. S. Baron and A. van Oudenaarden, "Unravelling cellular relationships during development and regeneration using genetic lineage tracing," *Nature Reviews Molecular Cell Biology* **20**, 753–765 (2019), doi:10.1038/s41580-019-0186-3.

[4] J. M. Kebschull and A. M. Zador, "Cellular barcoding: lineage tracing, screening and beyond," *Nature Methods* **15**, 871–879 (2018), doi:10.1038/s41592-018-0185-x.

[5] A. McKenna, G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier and J. Shendure, "Whole-organism lineage tracing by combinatorial and cumulative genome editing," *Science* **353**, aaf7907 (2016), doi:10.1126/science.aaf7907.

[6] B. Spanjaard, B. Hu, N. Mitic, P. Olivares-Chauvet, S. Janjuha, N. Ninov and J. P. Junker, "Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars," *Nature Biotechnology* **36**, 469–473 (2018), doi:10.1038/nbt.4124.

[7] B. A. Biddy, W. Kong, K. Kamimoto, C. Guo, S. E. Waye, T. Sun and S. A. Morris, "Single-cell mapping of lineage and identity in direct reprogramming," *Nature* **564**, 219–224 (2018), doi:10.1038/s41586-018-0744-4.

[8] B. Raj, J. A. Gagnon and A. F. Schier, "Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR–Cas9 barcodes by scGESTALT," *Nature Protocols* **13**, 2685–2713 (2018), doi:10.1038/s41596-018-0058-x.

[9] C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky and A. M. Klein, "Fundamental limits on dynamic inference from single-cell snapshots," *Proceedings of the National Academy of Sciences USA* **115**, E2467–E2476 (2018), doi:10.1073/pnas.1714723115.

[10] C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo and A. M. Klein, "Lineage tracing on transcriptional landscapes links state to fate during differentiation," *Science* **367**, eaaw3381 (2020), doi:10.1126/science.aaw3381.

[11] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, S. Gould, S. Liu, S. Lin, P. Berube *et al.*, "Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming," *Cell* **176**, 928–943.e22 (2019), doi:10.1016/j.cell.2019.01.006.

[12] A. Forrow and G. Schiebinger, "LineageOT is a unified framework for lineage tracing and trajectory inference," *Nature Communications* **12**, 4940 (2021), doi:10.1038/s41467-021-25133-1.

[13] S.-W. Wang, M. J. Herriges, K. Hurley, D. N. Kotton and A. M. Klein, "CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information," *Nature Biotechnology* **40**, 1066–1074 (2022), doi:10.1038/s41587-022-01209-1.

[14] W. Fang, C. M. Bell, A. Sapirstein, S. Asami, K. Leeper, D. J. Zack, H. Ji and R. Kalhor, "Quantitative fate mapping: A general framework for analyzing progenitor state dynamics via retrospective lineage barcoding," *Cell* **185**, 4604–4620.e32 (2022), doi:10.1016/j.cell.2022.10.028.

[15] J. S. Schiffman, A. R. D'Avino, T. Prieto, Y. Pang, Y. Fan, S. Rajagopalan, C. Potenski, T. Hara, M. L. Suvà, C. Gawad and D. A. Landau, "Defining heritability, plasticity, and transition dynamics of cellular phenotypes in somatic evolution," *Nature Genetics* **56**, 2174–2184 (2024), doi:10.1038/s41588-024-01920-6.

[16] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov *et al.*, "RNA velocity of single cells," *Nature* **560**, 494–498 (2018), doi:10.1038/s41586-018-0414-6.

[17] V. Bergen, M. Lange, S. Peidli, F. A. Wolf and F. J. Theis, "Generalizing RNA velocity to transient cell states through dynamical modeling," *Nature Biotechnology* **38**, 1408–1414 (2020), doi:10.1038/s41587-020-0591-3.

[18] M. Lange, V. Bergen, M. Klein, M. Setty, B. Reuter, M. Bakhti *et al.*, "CellRank for directed single-cell fate mapping," *Nature Methods* **19**, 159–170 (2022), doi:10.1038/s41592-021-01346-6.

[19] J. S. Herman, Sagar and D. Grün, "FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data," *Nature Methods* **15**, 379–386 (2018), doi:10.1038/nmeth.4662.

[20] S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler *et al.*, "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance," *Nature* **546**, 431–435 (2017), doi:10.1038/nature22794.

[21] B. L. Emert, C. J. Cote, E. A. Torre, I. P. Dardani, C. L. Jiang, N. Jain, S. M. Shaffer and A. Raj, "Variability within rare cell states enables multiple paths toward drug resistance," *Nature Biotechnology* **39**, 865–876 (2021), doi:10.1038/s41587-021-00837-3.

[22] J. Dai, S. Zheng, M. M. Falco, J. Bao, J. Eriksson, S. Pikkusaari *et al.*, "Tracing back primed resistance in cancer via sister cells," *Nature Communications* **15**, 1158 (2024), doi:10.1038/s41467-024-45478-7.

[23] Y. Oren, M. Tsabar, M. S. Cuoco, L. Amir-Zilberstein, H. F. Cabanos, J.-C. Hütter *et al.*, "Cycling cancer persister cells arise from lineages with distinct programs," *Nature* **596**, 576–582 (2021), doi:10.1038/s41586-021-03796-6.

[24] E. Traer, N. Javidi-Sharifi, A. Agarwal, J. Dunlap, I. English, J. Martinez, J. W. Tyner, M. Wong and B. J. Druker, "Ponatinib overcomes FGF2-mediated resistance in CML patients without kinase domain mutations," *Blood* **123**, 1516–1524 (2014), doi:10.1182/blood-2013-07-518381.

[25] S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran *et al.*, "A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations," *Cell* **141**, 69–80 (2010), doi:10.1016/j.cell.2010.02.027.

[26] J.-C. Marine, S.-J. Dawson and M. A. Dawson, "Non-genetic mechanisms of therapeutic resistance in cancer," *Nature Reviews Cancer* **20**, 743–756 (2020), doi:10.1038/s41568-020-00302-4.

[27] R. Salgia and P. Kulkarni, "The genetic/non-genetic duality of drug resistance in cancer," *Trends in Cancer* **4**, 110–118 (2018), doi:10.1016/j.trecan.2018.01.001.

[28] M. J. Hangauer, V. S. Viswanathan, M. J. Ryan, D. Bole, J. K. Eaton, A. Matov *et al.*, "Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition," *Nature* **551**, 247–250 (2017), doi:10.1038/nature24297.

[29] G. S. Kinker, A. C. Greenwald, R. Tal, Z. Ortigosa, C. J. Wagner, M. B. Regev *et al.*, "Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity," *Nature Genetics* **52**, 1208–1218 (2020), doi:10.1038/s41588-020-00726-6.

[30] A. Maynard, C. E. McCoach, J. K. Rotow, L. Harris, F. S. Haderk, J. M. Kerr *et al.*, "Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing," *Cell* **182**, 1232–1251.e22 (2020), doi:10.1016/j.cell.2020.07.017.

[31] C. Bunne, S. G. Stark, G. Gut, J. Sarabia del Castillo, M. Levesque, K.-V. Lehmann, L. Pelkmans, A. Krause and G. Rätsch, "Learning single-cell perturbation responses using neural optimal transport," *Nature Methods* **20**, 1759–1768 (2023), doi:10.1038/s41592-023-01969-x.

[32] A. E. Rodriguez-Fraticelli and V. Parreno, "Charting single-cell lineages with synthetic and natural barcodes," *Nature Reviews Genetics* **27**, 442–461 (2026), doi:10.1038/s41576-026-00943-5.

[33] K. Wang, X. He and Z. Hu, "Computational approaches for multimodal lineage tracing," *Nature Reviews Genetics* **27**, 779–796 (2026), doi:10.1038/s41576-026-00969-9.

[34] K. Jindal, M. T. Adil, G. C. Rivera-Gonzalez *et al.*, "Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes," *Nature Biotechnology* **42**, 946–959 (2024), doi:10.1038/s41587-023-01931-4.