



Published in final edited form as:

Cell. 2022 November 23; 185(24): 4604–4620.e32. doi:10.1016/j.cell.2022.10.028.

Quantitative fate mapping: A general framework for analyzing progenitor state dynamics via retrospective lineage barcoding

Weixiang Fang^{1,2,3}, Claire M. Bell^{4,5}, Abel Sapirstein^{1,3,6}, Soichiro Asami^{1,3}, Kathleen Leeper^{1,3}, Donald J. Zack^{5,7,8,9}, Hongkai Ji^{2,*}, Reza Kalhor^{1,3,4,7,8,9,10,*}

¹Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

²Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

³Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

⁴Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

⁵Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

⁶H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

⁷Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

⁸Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

⁹Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

¹⁰Lead contact

Summary

*Correspondence: kalhor@jhu.edu (R.K.), hji@jhu.edu (H.J.).

Author Contributions

W.F. and R.K. conceived the project and designed the experiments; W.F. established statistical frameworks, carried out computational analyses; H.J. provided statistical advice; C.M.B. created and characterized cell lines and carried out in vitro experiments with assistance from W.F. and K.L.; A.S. helped create Phylotime; S.A. created the inDelphi-based allele predictions; R.K. and W.F. interpreted the data and wrote the manuscript with assistance from C.M.B. and A.S. and input from H.J.; D.J.Z, H.J., and R.K. supervised the project.

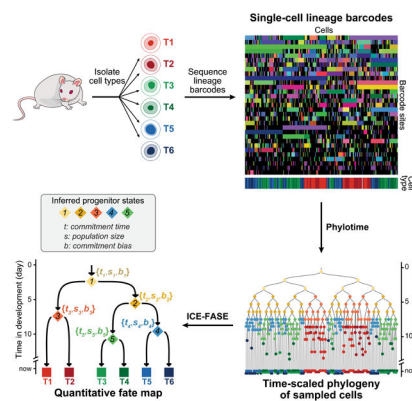
Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interests

The authors declare no competing financial interests.

Natural and induced somatic mutations that accumulate in the genome during development record the phylogenetic relationships of cells; whether these lineage barcodes capture the complex dynamics of progenitor states remains unclear. We introduce quantitative fate mapping, an approach to reconstruct the hierarchy, commitment times, population sizes, and commitment biases of intermediate progenitor states during development based on a time-scaled phylogeny of their descendants. To reconstruct time-scaled phylogenies from lineage barcodes, we introduce Phylotime, a scalable maximum likelihood clustering approach based on a general barcoding mutagenesis model. We validate these approaches using realistic *in silico* and *in vitro* barcoding experiments. We further establish criteria for the number of cells that must be analyzed for robust quantitative fate mapping and a progenitor state coverage statistic to assess the robustness. This work demonstrates how lineage barcodes, natural or synthetic, enable analyzing progenitor fate and dynamics long after embryonic development in any organism.

Graphical Abstract



In Brief:

Estimating the temporal distance between cells using lineage barcodes enables reconstructing the hierarchy, commitment times, population sizes, and commitment biases of their progenitor states during development.

Keywords

lineage tracing; quantitative fate mapping (QFM); progenitor state dynamics; progenitor field; somatic mutations; Phylotime; ICE-FASE; coalescent theory; time-scaled cell phylogeny

Introduction

Embryonic development is the genesis of complex body plans in the animal kingdom. It starts with the zygote, a single cell in a totipotent state, and ends with thousands of specialized terminal cell types organized in tissues. In between, dividing cells traverse a hierarchy of increasingly diverse but decreasingly potent intermediate progenitor states. Each progenitor state specifies the ensuing states that its descendant cells may assume, thus directing their fates. Collectively, progenitor states orchestrate the emergence of terminal

cell types to form complex tissues. Therefore, delineating how progenitor states specify cell fate is critical for understanding normal and dysregulated development.

The recent advances in genome engineering and sequencing have inspired a new approach for interrogating cell fate: retrospective lineage analysis using synthetic or natural somatic DNA barcodes. These approaches rely on the accumulation of random mutations in the genome during development. Each mutation is inherited by the descendants of the cell in which it occurs; each descendant can add new mutations to the combination it inherited. This process marks each cell with a barcode—a combination of mutations—that encodes its phylogenetic relationship to the other cells¹. Synthetic lineage barcoding, which relies on gene editing technologies to induce mutations, has been implemented in model organisms such as zebrafish^{2–4} and mouse^{5–7}. Natural lineage barcoding, which relies on naturally-occurring somatic mutations, has been primarily used in humans^{8,9}. These retrospective approaches hold a unique promise for mapping cell fate. Unlike prospective lineage tracing approaches¹⁰, they have the potential to resolve entire hierarchies of progenitor states, thereby facilitating the analysis of non-cell autonomous effects. Unlike single-cell molecular profiling approaches, they can bridge time gaps between terminal cells and their progenitors that existed far earlier in time. Moreover, they can be applied to humans and non-model organisms where analyzing somatic mutations from cadavers can be more practical and ethical than analyzing embryos.

Despite this compelling potential, the full scope of the information that lineage barcoding can provide about the fate of the intermediate progenitor states remains unclear for multiple reasons. First, cell phylogeny is a function of cell divisions and most cell divisions in higher organisms do not accompany fate decisions. In the roundworm *C. elegans*, a unique model organism in which almost all cell divisions give rise to daughters with different fates, the phylogeny of terminal cells is identical to the fate of their progenitors¹¹. However, in more complex organisms, progenitor populations can undergo cell divisions that are not associated with fate decisions, leading to divergences between phylogeny and fate^{12,13}. As a result, liver hepatocytes of an identical progenitor state history may have the maximum possible distance on the phylogenetic tree by being the descendants of different cells at the 2-cell stage. Second, while the progenitor states and their fate remain largely stereotyped within species, the phylogenetic histories of the cell populations that assume those progenitor states can vary greatly from embryo to embryo due to stochasticity in fate decisions^{14,15}. As a result, phylogenies of different subsets of cells from different embryos cannot be combined to synthesize a full picture the same way they can be in *C. elegans*. Third, single-cell lineage barcodes can be obtained for only a small sample of cells as current technologies can only sequence thousands of single cells whereas most mammals have millions of cells in each tissue. Given the divergences between fate and phylogeny and the variable nature of the latter, it remains unclear how phylogenies derived from small samples can reliably inform organism-level fate maps. Complicating matters further, phylogenetic inference from lineage barcodes is inherently subject to error because a finite number of barcoding sites may not record every cell division¹⁶, and even with infinite barcodes, finding the optimal tree is still a computationally intractable (NP-hard) problem¹⁷. Collectively, these considerations raise critical questions about the value of measuring cell phylogeny through barcoding approaches

in complex organisms: What features of progenitor states are reflected in the phylogeny of a limited sample of cells? How can these features be extracted from lineage barcodes?

To address these questions, we systematically studied the relationship between cell fate and cell phylogeny as derived using lineage barcodes. First, we established a method to generate cell phylogenies and developed the ICE-FASE algorithm to reconstruct quantitative fate maps—models that represent the hierarchy and dynamics of progenitor states—from time-scaled phylogenies. We found that quantitative fate mapping requires adequate representation of each progenitor state’s progeny among the sampled terminal cells. Second, we simulated synthetic lineage barcoding in mice and established Phylotime, a general and scalable method to infer time-scaled phylogenies from lineage barcodes. We found that Phylotime-inferred phylogenies enable robust quantitative fate map reconstruction when coupled with ICE-FASE. Overall, this work establishes quantitative fate mapping as a general framework for capturing the fate hierarchy and dynamics of progenitor populations using lineage barcodes of their descendants.

Results

Quantitative fate map: a model of progenitor field dynamics in development

We began by establishing the quantitative fate map, a model of the progenitor state hierarchy that gives rise to a group of cell types (Figure 1A). Each progenitor state is defined by its potency, which is the set of cell types it can produce. It is also associated with a commitment event, when its cells transition to less potent downstream states. The commitment event confers each progenitor state three additional defining features: i) commitment time, which is the time when a progenitor state’s cells commit to its downstream states, ii) population size, which is its number of cells at commitment time, and iii) commitment bias, which is the proportions of its population committing to each downstream state (Figure 1B). Commitment times represent the order and pace of developmental events¹⁸, population sizes are important determinants of organ and tissue size¹⁹, and commitment biases reflect the epigenetic state of progenitors and the effect of non-cell autonomous cues²⁰. The progenitor states are unobserved in barcoding experiments; only their descendants at the time of sample collection are observed. We refer to these observed descendants as terminal cell types. In summary, a quantitative fate map defines the fate dynamics of a progenitor field²¹—a collection of progenitor states that give rise to a set of observed cells.

A diverse test panel of quantitative fate maps

We constructed 331 quantitative fate maps covering diverse developmental scenarios (Figures 1C and S1A, **Methods**). Representing increasing field sizes, the maps are in three categories of 16, 32 or 64 terminal cell types. We label progenitor states and terminal types with “P”s and “T”s followed by numerals, respectively. Within each category, the topologies of the maps range from perfectly balanced to highly unbalanced (Figures 1C and S1B,C) as measured by the Colless imbalance index²². In more unbalanced maps, progenitor states split into increasingly unequal diversities of terminal types (Figure 1C). In addition to size and topology, the parameters of progenitor states within each map vary (Figures S1D–J). Commitment times are between $t = 2.5$ and 10.9 days, which roughly correspond

to the beginning of fate restrictions and the end of organogenesis in mouse development (Figures 1A–C and S1D). Commitment biases were drawn from a beta distribution and cover a wide range (Figures 1B and S1E, **Methods**). Cell division and death rates were drawn from uniform distributions, ranging from 0.6 to 0.35 days per doubling and 0.02 to 0.08 death probability per division, respectively (Figures S1F–H), broadly matching reported rates during mouse embryogenesis^{23,24}. As all fate maps start with one founder cell at time $t = 0$, the division and death rates together with other fate map parameters dictate the progenitor population sizes at each point in time (Figures 1B and S1I,J). All fate maps end at $t = 11.5$ days when terminal cell types are sampled for observation based on either fixed or proportional sampling (Figure 1B). Under fixed sampling, the same number of cells are sampled from each terminal type, imitating experiments where target terminal cell types are purified using sorting or other methodology. Under proportional sampling, each terminal type is sampled based on its share of the total population, imitating experiments where cells are sampled without enrichment from whole tissues.

Modeling cell phylogeny based on a quantitative fate map

We next established a generative model to simulate cell phylogenies based on each quantitative fate map. Generating the entire tree of cell divisions for millions of cells (Figure 1B) is computationally impractical. To overcome this problem, we developed a model based on coalescent theory in population genetics^{25–27} to generate time-scaled phylogenies only for sampled terminal cells rather than for all terminal cells (Figure S2, **Methods**). In brief, after choosing the number of cells to be sampled from each terminal cell type of a fate map, the number of cells from each progenitor state ancestral to sampled terminal cells is generated at all prior time points going backward to the founder cell. These sampled terminal and progenitor cells compose the nodes of the phylogenetic tree. Edges are then assigned by randomly connecting nodes from each time point to their progenitor nodes in the earlier time point. This approach generates time-scaled phylogenies for sampled cells based on their progenitors' fate map (Figure 1D) in a computationally efficient manner.

Using this model, we generated time-scaled phylogenies for all quantitative fate maps in our test panel, sampling an average of 100 cells per terminal type under both fixed and proportional sampling (Figure 1B). To capture the variable nature of phylogeny, we simulated five phylogenies for each condition, representing a different set of cells being sampled from the same set of terminal cell types in different individuals (Figure S2). Together, these results represent $3,310$ experiments (331 maps \times 2 sampling schemes \times 5 repeats) wherein phylogeny is known for a small fraction of cells (average 0.07%) derived from a complex field of progenitors. We will use these phylogenies to benchmark fate mapping algorithms established below.

Reconstructing the hierarchy of progenitor states from cell phylogeny

To derive fate map topology from time-scaled phylogenies, we used the timings of apparent fate separations between terminal cell types. First, we annotated each node in the phylogenetic tree with its observed fate—the types of its observed terminal descendants (Figure 2A). Next, we identified the nodes whose observed fates are more potent than that of both their daughter nodes (Figures 2A,B). For instance, if an internal node leads to terminal

cell types {T3, T4, T5} but {T3, T4} are only seen in one of its branches and {T5} only in the other, this node constitutes a FASE (FAte SEparation) between T3 and T5 as well as between T4 and T5. The average time since FASEs between two terminal cell types (i.e., FASE distance) measures their developmental distance: long FASE distances indicate early separation in development and short FASE distances indicate more recent separation (Figures 2B,C). We can thus compile a matrix of FASE distances between all terminal cell types (Figure 2D) and apply a clustering method (UPGMA) to obtain fate map topology (Figure 2D, **Methods**). This fate map topology establishes a hierarchy of increasingly diverse but decreasingly potent “inferred” progenitor states (labeled with “*iP*”s followed by numerals) that give rise to the observed cell types (Figure 2E). To summarize, the FASE algorithm reconstructs a hierarchy of inferred progenitor states based on the patterns of potency restriction in the phylogeny.

We applied the FASE algorithm to reconstruct fate map topology for each simulated phylogeny in our panel of 3,310 (Figure S1K). For comparison, we also used the shared progenitor score (SPS) which estimates the distance between two terminal cell types based on the number of nodes in the phylogenetic tree that have those terminal types in their observed fate, weighted by how many other terminal types are among the observed fates⁶. We further generated random fate map topologies as negative control (**Methods**). In all cases, we compared the reconstructed topology to that of its corresponding true fate map using the Kendall-Colijn (KC) distance with its tuning parameter (λ) set to zero (KC_0)²⁸. The KC distance measures the difference between rooted trees. It compares the placement of the most recent common ancestor (MRCA) of all pairs of tips relative to the root based on either the number of edges (KC_0) or path length (KC_1). A KC_0 distance of zero indicates that the reconstructed and true fate maps have identical topologies (a KC_1 distance of zero, which we will use later, indicates identical topologies and branch lengths between two trees). The results show that FASE strategy consistently outperforms SPS (Figure 2F): It predicts perfectly accurate topologies when the fate map is small or has low imbalance; it predicts informative topologies even for large fate maps with extreme imbalances. Unbalanced maps have smaller intervals between commitment events (Figure S1L), making reconstruction of topology more challenging. We also observed that fixed sampling outperforms proportional sampling (Figure 2F), likely because it ensures better representation of rare terminal populations. These results establish the FASE algorithm as a robust and scalable method to reconstruct fate map topology from cell phylogeny.

To better understand the sources of topology reconstruction inaccuracies, we investigated FASE distance error, which is the difference between estimated FASE distance between two terminal cell types and the time since their last common progenitor state in the true fate map. We found a negative correlation between FASE distance error and the sampling fractions of progenitor states (Pearson’s $R=-0.40$, $p < 2.2e-16$) (Figure S3A) in the 3,310 simulated experiments. Sampling fraction is the proportion of the progenitor state population whose progeny are represented among sampled terminal cells and is known from simulations (Figure 1B). Later progenitor states tend to have larger population sizes and thus lower sampling fractions (Figure S3B). Undersampling the terminal descendants of a progenitor cell can bias its corresponding FASE distances (Figure S3C). To validate

the effect of sampling fraction, we generated phylogenies for all our 16-terminal type fate maps, sampling an average of 25, 50, 100, and 200 cells per terminal type. We observed that increased sampling reduces topology reconstruction error (Figure S3D). For example, every doubling of the number of sampled cells, which increased the sampling fraction of progenitor states by 33% on average, increased the percent of perfectly reconstructed topologies by 18.9% on average. These results establish progenitor state sampling fraction as an important parameter for fate map topology reconstruction.

Estimating progenitor state commitment time from cell phylogeny

To characterize the dynamics of inferred progenitor states, we turned to the internal nodes of the time-scaled phylogenies. We assigned each internal node in the phylogenetic tree to the least potent inferred progenitor state or terminal type from the reconstructed fate map topology that contained the node's observed fate (Figure 3A). For example, a node with an observed fate of {T2, T3, T4} can be assigned a more potent inferred state of *iP2* capable of {T1 to T5} if the now-reconstructed fate map topology (Figure 2E) indicates that *iP2* differentiates into fates {T1, T2} and {T3, T4, T5} (Figure 3B). To assess the fidelity of these assignments, we compared the inferred states of internal nodes in all 3,310 phylogenies to their true states which are known from simulations (Figure S3E). The only type of error was assigning an internal node to a progenitor state less potent than its true state, which occurred, on average, for 27.7% of the assignments in each phylogeny (Figures S3E,F). This type of error is also caused by undersampling (Figure S3G); hence, we will keep track of progenitor states' sampling fractions going forward.

To derive the commitment time of each progenitor state, we defined Inferred Commitment Events (ICEs): An ICE is a node whose inferred state is more potent than that of both of its immediate descendants (Figures 3B,C). For example, in Figure 3B, when an internal node assigned to *iP2* (capable of {T1 to T5}) splits into two nodes with assigned states of *T2* and *iP14* (capable of {T3, T4}) respectively, we count this node as an ICE for *iP2*. ICEs improve on FASEs by leveraging the now-reconstructed fate map topology to identify a more confident set of nodes that represent state transitions. We defined the commitment time for a progenitor state as the mean of its ICE times (Figure 3D). Across all progenitor states in our panel of 3,310, ICE times captured the relative timing of commitment events as indicated by a high rank correlation (Spearman's $\rho=0.90$ for fixed sampling and 0.91 for proportional sampling) (Figure 3E). Like FASE distance error, the error of estimated commitment time for progenitor states showed a strong negative correlation with their sampling fraction (Pearson's $R = -0.97$) (Figure 3F). In fact, when only considering progenitor states with a sampling fraction above 0.25 (Figure 3F), ICE times not only captured relative commitment times almost perfectly (Spearman's $\rho=1.00$ for fixed and 0.99 for proportional sampling) but also captured the exact timing of commitments as indicated by a low root mean square error (RMSE=0.31 days for fixed and 0.27 days for proportional sampling) (Figure 3E). These results establish ICE times as estimates for the commitment times of the progenitor states from time-scaled phylogenies of sampled cells. They also demonstrate the central importance of progenitor state sampling fraction for obtaining reliable estimates.

Estimating progenitor state population size and commitment bias from cell phylogeny

We next leveraged both fate map topology and commitment times to estimate population sizes and commitment biases of progenitor states. To estimate population size for a progenitor state, we identified the subset of all edges in the phylogeny that (i) cross the progenitor state's commitment time and (ii) connect nodes assigned as either the progenitor state itself or any of its upstream or downstream states in fate map topology (Figures 3G,H, **Methods**). These branches represent cells of the progenitor state that are present at its estimated commitment time. We thus counted the number of incoming nodes to these branches as the population size (Figures 3H,I). For commitment bias, we calculated the proportion of these branches that end in each of the downstream fates irrespective of their parental state (Figures 3H,J). Applying this algorithm to the 3,310 simulated time-scaled phylogenies, we found that the ability to estimate population size and commitment bias for a progenitor state depend heavily on its sampling fraction as well as the sampling method. For well-sampled progenitor states (sampling fraction > 0.25), population size estimates agree well with their true sizes (Spearman's $\rho=0.98$ for fixed and $\rho=0.90$ for proportional sampling) (Figure 3K). For undersampled progenitor states, a progenitor population's size estimate is capped at the number of its sampled terminal progeny, which is reasonably informative in proportional sampling (Spearman's $\rho=0.78$) but uninformative in fixed sampling (Spearman's $\rho=0.06$) (Figure 3K). For commitment bias, proportional sampling produced good estimates for the well-sampled progenitor states (Spearman's $\rho=0.83$) and reasonable estimates for the undersampled ones (Spearman's $\rho=0.72$). (Figure 3L). Fixed sampling produced good estimates only for well-sampled progenitor states (Spearman's $\rho=0.82$) and was almost uninformative for undersampled ones (Spearman's $\rho=0.15$) (Figure 3L). Consistent with these results, average population size and commitment bias estimate errors approach zero as sampling fraction approaches one (Figures S3H,I). The more effective estimation of population size and commitment bias with proportional sampling is due to the inherent correlation between the size of terminal populations, which proportional sampling captures, and that of their progenitors. Taken together, these results establish a strategy for estimating progenitor population size and commitment bias from a time-scaled phylogeny of cells.

Modeling realistic lineage barcoding results in development

So far, we have established a strategy—the ICE-FASE algorithm—to reconstruct quantitative fate maps which describe progenitor state hierarchy, commitment time, population size, and commitment bias from time-scaled phylogenies. The phylogenies that were used thus far represent the exact sequence and timing of cell divisions as simulated (i.e., true phylogeny). In actual experiments, phylogeny must be inferred from lineage barcodes and such inferences are inherently subject to error due to the limitations of both barcoding systems and computational techniques. Therefore, how close any inferred phylogeny is to the true phylogeny remains uncertain. To address whether quantitative fate maps can be obtained from inferred phylogenetic trees despite their uncertainty, we started by simulating realistic lineage barcoding outcomes. To do so, we established a mutagenesis model comprising independent barcoding sites that accumulate mutations according to a Poisson point process with a constant rate (Figures 4A,B). Each mutation converts an active

copy of the site into one of many possible inactive mutated alleles, each with a distinct emergence probability. Next, we set the parameters of this model based on the MARC1 system⁵ wherein barcoding sites are homing guide RNA loci (hgRNAs)²⁹. We estimated the mutation rates of MARC1 hgRNA from published embryonic time course measurements³⁰ (Figure S4A, Table S1, **Methods**). We estimated emergence probabilities of mutant alleles for each hgRNA by modifying the inDelphi machine learning algorithm³¹, which predicts CRISPR-Cas9 mutations (Figure S4B, **Methods**). We validated the mutation model and its parameters by simulating barcoding in whole-mouse embryos and comparing the results to that of actual MARC1 barcoded embryos (**Methods**). The comparison showed a broad agreement between experimental and simulated barcoding with respect to the emergence probability of mutant alleles, total mutation levels over the course of embryogenesis, and the diversity and composition of mutant alleles within an embryo (Figures S4B–G). These results suggest that our strategy produces realistic barcoding results that are comparable to lineage barcoding in mouse embryos. We thus simulated mutagenesis in our panel of 3,310 phylogenies with 50 hgRNAs per cell (**Methods**). These simulations yielded *in silico* barcoding experiments wherein, similar to actual experiments, the barcodes and terminal types are known for sampled single cells (Figure 4C).

Inferring time-scaled cell phylogenies from single-cell lineage barcodes using Phylotime

For quantitative fate mapping, the phylogenetic tree inferred from lineage barcodes must be time-scaled. However, many inference methods lack a mutagenesis model specific to lineage barcoding and thus their resulting phylogram branch lengths do not represent interdivision times. Those with a barcoding mutagenesis model^{32–34} require optimization techniques that do not scale to thousands of cells, as sampled here in each simulated experiment. Therefore, we developed a scalable method to infer time-scaled phylogenies from lineage barcodes. In this method, we first compute a maximum likelihood estimate of the time that separates a pair of cells from their most recent common ancestor (time since MRCA) for all pairs of terminal cells (Figures 4D and S4H, **Methods**). We then apply UPGMA hierarchical clustering to the pairwise time since MRCA matrix to obtain a time-scaled phylogenetic tree (Figure 4E). We call this approach, which scales in polynomial time, PHYlogeny inference using Likelihood Of TIME (Phylotime).

To evaluate Phylotime's performance, we first compared estimated times since MRCA for all pairs of cells from a simulated barcoding experiment to those derived from the corresponding true phylogeny (Figure S4I) and found that the two were highly correlated (Pearson's $R = 0.93$). We then simulated barcoding with 25, 50, and 100 hgRNAs in all 530 simulated phylogenies for the 16-terminal type fate maps and applied Phylotime to infer phylogeny from each. For comparison, we also inferred phylogenies using a Hamming distance-based clustering method as well as Cassiopeia³⁵ which is a heuristic approach based on maximum parsimony (**Methods**). Other common methods do not scale well to this number of terminal cells (1,600) and barcoding sites (up to 100) in each experiment. We then evaluated the difference between inferred phylogenies and their corresponding true phylogenies using KC_0 distance for topology and KC_1 distance for combined topology and branch length (Figures 4F,G). KC_0 results showed that Phylotime produced topologies that had on average 80% less error compared to Cassiopeia and 6% less error compared to

Hamming (Figure 4F). KC_1 results showed that Phylotime's time-scaled phylogenies had on average 82% less error compared to Cassiopeia and 78% less error compared to Hamming (Figure 4G). Importantly, only Phylotime's solutions converged to the true phylogeny with an increasing number of barcoding sites (Figure 4G). Together, these results show that Phylotime can accurately infer time-scaled phylogenies from lineage barcodes.

Reconstructing quantitative fate maps from lineage barcodes

With realistic barcoding simulation and a method to infer time-scaled phylogenies in hand, we finally assessed how well lineage barcodes can inform quantitative fate maps. We applied the ICE-FASE algorithm to all 3,310 Phylotime-inferred time-scaled phylogenies from experiments simulated with 50 hgRNAs to reconstruct quantitative fate maps and compared these fate maps to those obtained by applying ICE-FASE to the true phylogeny. Phylotime-inferred phylogenies performed almost as well as true phylogenies at reconstructing fate map topology, regardless of fate map size and imbalance or sampling strategy (Figure 5A), with 22% more error on average compared to the true phylogeny. Similarly, for commitment times, population sizes, and commitment biases of progenitor states, Phylotime-inferred phylogenies performed similarly to true phylogenies in all conditions, recovering on average 95%, 80%, and 76% of the correlation, respectively (Figure 5B). Taken together, these results show that quantitative fate maps can be faithfully reconstructed based on time-scaled phylogenies inferred using Phylotime despite errors inherent to phylogenetic inference.

Effect of barcoding site composition, experimental errors, and sampling time on quantitative fate mapping

We next analyzed how the number and the mutation rates of barcoding sites affect the ability to reconstruct quantitative fate maps from lineage barcodes. On the panel of 3,310 true phylogenies, we repeated barcoding simulation with 25, 50, or 100 hgRNAs per cell mutating at fast, intermediate, or slow rates, as observed in MARC1 hgRNAs (Figure S4A). We then applied Phylotime followed by ICE-FASE to these 29,790 simulated experiments to obtain quantitative fate maps. To assess these maps, we instituted a general evaluation strategy that will also be used in all later sections: we used KC_0 error for topology; we used the RMSEs of progenitor state commitment times, $\log_2(\text{population sizes})$, and commitment biases irrespective of sampling fraction and based on the true topology (using imperfectly reconstructed topology would bias estimates in favor of easily resolvable progenitor states) (**Methods**). Note that RMSE measures absolute error; experiments with relatively high RMSEs often still have accurate relative ordering of progenitor state parameters. The analysis showed that very slow rates of mutations (~ 0.002 mutations per site per division) cannot resolve reliable quantitative fate maps under the conditions tested here (Figure S4J). Though increasing the number of slow-mutating sites increased reconstruction accuracy generally, indicating that the combined mutation rates of all sites ultimately dictate reconstruction efficiency (Figure S4J). Intermediate rates of mutation (~ 0.05 per site) performed better than fast mutation rates (~ 0.28 per site). As fast sites are more likely to record information only in the earlier days of development (Figures S4D–F) and before many of the progenitor states have come into existence (Figure S1D). Overall, these results show that for quantitative fate mapping, an adequate overall level of mutagenesis must take place during the development window of progenitor states.

We next considered two common modes of experimental error: allele dropout and allele switching. We carried out simulations with increasing levels of error in the resulting data and assessed the fidelity of quantitative fate map reconstruction (**Methods**). For missing alleles, before applying Phylotime, we imputed the missing alleles for each barcoding site sequentially by leveraging information from all other sites (**Methods**). The results, which are detailed in Figures S5A–D, suggest that while the accuracy of estimates declines with increasing experimental error, quantitative fate mapping algorithms can tolerate errors and behave stably in response.

We also simulated sampling of terminal types one to five days after the last commitment event. The results, which are detailed in Figures S5E–M, show that a progenitor field may be assessed by sampling its terminal cell types or the descendants of those terminal cell types at any time after its development.

Cell death, non-random sampling, and asymmetric divisions can empower quantitative fate mapping

Because cell death is a prevalent developmental control mechanism³⁶, we carried out simulations with increasing rates of cell death in either all terminal types or all progenitor states. As detailed in Figures S5N–S, we found that cell death in progenitor states drastically improved estimates of commitment time, population size, and commitment bias by bottlenecking progenitor states to effectively increase their sampling fractions. These results in turn suggest that non-random sampling of terminal types can be employed to effectively bottleneck their progenitors and facilitate analyzing large progenitor states. Conversely, cell death in terminal states did not significantly alter reconstruction, suggesting that a progenitor state may be analyzed using a subset of its downstream states or types.

We also considered stereotyped cell fate commitment with asymmetric cell divisions, which are common in the development of the nervous system³⁷. We simulated barcoding experiments on a pectinate 16 terminal type fate map (Figure 1C, bottom left) with commitments happening exclusively through asymmetric divisions. The results, which are detailed in Figures S5T–Z, indicate that asymmetric divisions facilitate quantitative fate mapping.

Consensus quantitative fate maps from multiple biological replicates

We expanded the ICE-FASE algorithm to take advantage of multiple biological replicates assuming they have a common underlying fate map. Briefly, to obtain consensus topology from replicates, their FASE distance matrices were averaged before applying the clustering algorithm (Figures 2D,E). To obtain consensus commitment times, the consensus topology was used to assign inferred progenitor states to each replicate's phylogeny and ICE times of each progenitor state were pooled from different replicates (Figures 3A–D). To obtain consensus population sizes and commitment biases, the consensus commitment time for each progenitor state was used to estimate its population size and commitment bias in each phylogeny (Figures 3G–K) and these estimates were averaged to obtain respective consensus.

We then assessed fate map reconstruction accuracy based on the number of replicates. For each of our 53 fate maps with 16 terminal types (Figure 1C), we simulated 10 replicates of barcoding experiments each for fixed and proportional sampling with an average of 50, 100, or 200 sampled cells per terminal cell type (53 maps \times 2 sampling schemes \times 3 sample sizes = 318 sets of 10 replicates). For each replicate, we applied Phylotime to obtain a time-scaled phylogeny. We then randomly combined 1, 2, 3, 4, or 5 of the 10 replicates to reconstruct consensus quantitative fate maps and compared their accuracies. The results show that additional replicates indeed reduce error of all quantitative fate map estimates (Figures 5C–F). However, the gains with additional replicates are diminishing and do not converge to zero error. This observation is best explained by undersampled progenitor states whose estimates in each replicate can be uninformative (Figure 5B). In fact, estimation error for the well-sampled progenitor states approached zero with additional replicates while it did not for the undersampled ones (Figures 5G–I).

Importantly, the gains achieved by sequencing more cells per replicate outweighed those achieved by combining more replicates. For example, quadrupling the number of replicates from one to four when sampling 50 cells decreased average error in topology, commitment time, population size, and commitment bias by 30%, 23%, 4%, and 3% respectively (Figures 5C–F). By contrast, a quadrupling of the number of cells sequenced in a single replicate from 50 to 200 decreased average error in topology, commitment time, population size, and commitment bias by 46%, 41%, 26%, and 12% respectively (Figures 5C–F). These results suggest that limited resources can be better spent on sequencing more cells per individual than a larger number of individuals.

Resolving multifurcations and prolonged commitments

Because some progenitor states undergo gradual commitment during development, we simulated a progenitor state committing over a 0.5-, 1.1-, or 1.7-day commitment window. The results, which are detailed in Figures S6A–E, indicate that the ICE-FASE algorithm can resolve progenitor states that gradually commit over a window of time.

Because progenitor states in development may simultaneously commit to more than two immediate downstream states or types, we simulated barcoding in a fate map where a progenitor state undergoes trifurcation. The results, which are detailed in Figures S6F–L, show that the ICE-FASE algorithm can resolve multifurcations in fate maps given adequate sampling.

Assessing robustness of quantitative fate map estimates from experimental data with PScov

Despite its central importance in evaluating the robustness of quantitative fate map reconstruction, progenitor state sampling fraction cannot be directly obtained from sampled terminal cells or their phylogeny. To address this gap, we introduce estimated progenitor state coverage (PScov), a proxy for sampling fraction that can be derived from sampled cells alone. PScov is defined as the number of observed terminal progeny for a progenitor state divided by its estimated population size (Figure 6A). Intuitively, this statistic indicates how many terminal descendants were sampled per inferred progenitor cell. We found that a high

PScov is indeed predictive of a high sampling fraction (Figure 6B). For example, in our 3,310 simulated experiments, 69.8% of all progenitor states that are sampled more than 25% also have a PScov larger than 5, and more than 99.9% of progenitor states with PScov larger than 5 have sampling fractions larger than 25% (Figure 6C). Moreover, average commitment time, population size, and commitment bias errors approach zero with increasing PScov (Figures 6D–F). Therefore, PScov makes it possible to assess the robustness of quantitative fate map parameters for each progenitor state in actual experiments.

Experimental validation of quantitative fate mapping

To test this quantitative fate mapping framework, we used an experimental system in which quantitative fate map parameters can be known. We established a human induced pluripotent stem cell (iPSC) line with 32 hgRNA barcoding sites (Table S2, **Methods**). The line also includes inducible Cas9 to activate barcoding. 24 of the 32 hgRNAs were determined to be active and accumulated random mutations upon Cas9 induction (Figure S7A). We then designed growing and splitting schemes in culture that mimic progenitor state hierarchies (Figure 7A). In two parallel experiments, starting from single cells, we initiated barcoding and passaged growing cells into subpopulations at known times, numbers, and split ratios (Figures 7B and S7B,C). The two experiments were similar except that in one (E1), progenitor state 3 (P3) was split two days before progenitor state 4 (P4), whereas in the other (E2), P4 was split two days before P3. In effect, the last populations of split cells in these experiments represent terminal cell types and their ancestral populations represent progenitor states. In the end, we sequenced barcodes from 192 single cells in each terminal population (**Methods**). After data processing, we obtained on average 158 cells per terminal population with a median of 26 hgRNAs detected per cell (Figures 7C and S7D,E). We also conducted simulations on the E1 and E2 reference fate maps (Figure S7B) with cell division rates derived from the actual progenitor population sizes at each split and hgRNA mutation rates (Figure S7C) and allele emergence probabilities respectively obtained from time-course measurements and inDelphi predictions (Figure S7F, Table S2). We applied Phylotime and ICE-FASE to both simulated and experimental data. The experimental data reconstructed the topology correctly for both E1 and E2 (Figure 7D), and so we refer to the inferred states by their true state names hereafter. In addition to the correct topology, the inferred fate maps recovered the correct orders of commitment in both experiments (Figure 7D), including the relative commitment times for P3 and P4 which were switched between E1 and E2. This result suggests that our strategy can identify quantitative fate map differences in different systems.

PScov ranged from 1.68 to 2.36 in all progenitor states other than the founder (P5), indicating that they were highly undersampled. As such, we did not expect to recover the exact commitment times and population sizes of the passaged intermediate populations. Nevertheless, we evaluated if these estimates would approach the truth with increased sampling. We repeated quantitative fate map reconstruction 50 times each with varying numbers of terminal cells subsampled from the total sequenced population. In parallel, we carried out simulations with the same subsample sizes on reference fate maps. We then classified inferred fate maps based on their topology and correctness of relative ordering (Figure S7G) and found that the fraction of correct topology and relative order

of commitment increased in a similar fashion with increasing sampling in both the simulated and experimental datasets. Additionally, commitment time and population size estimates for P3 and P4 approached their true values with increasing sampling in simulated and experimental sets alike (Figures S7H,I). Together, these observations validate our barcoding models used for simulation and suggest that our quantitative fate mapping strategy, ICE-FASE, and Phylotime are robust to natural variations in cellular behavior.

Discussion

In this study, we have established a robust and versatile approach, called quantitative fate mapping, to use cells' lineage barcode and identity information to retrospectively characterize the progenitor field that gave rise to them. Quantitative fate mapping involves two stages. The first stage entails inferring a time-scaled phylogenetic tree from single-cell lineage barcodes. To do so, we have developed Phylotime, which scales to large trees with thousands of terminal branches. Moreover, its mutagenesis model can be adapted to other systems with multiple independent mutation sites such as natural somatic mutations. The second stage entails reconstructing the hierarchy of progenitor states that led to the observed cell types and estimating their commitment times, population sizes, and commitment biases. To do so, we have developed the ICE-FASE algorithm which uses nodes in the time-scaled phylogenetic tree that are associated with fate decisions as chronometers of progenitor population dynamics. While other studies have focused on reversible cell state transitions³⁸, non-cumulative barcoding^{15,39}, or single progenitor states^{34,40}, our approach is unique in that it evaluates the dynamics of progenitor states using cumulative somatic mutations, scales to large and complex progenitor fields, and can tolerate the errors that are inherent to phylogenetic inference.

The choice of terminal cell types decides the progenitor states that are analyzed in a quantitative fate map. As a simplified example, taking ectoderm-derived neurons, mesoderm-derived myocytes, and endoderm-derived hepatocytes in adult mice as terminal cell types would analyze their common progenitor states prior to gastrulation. Therefore, quantitative fate mapping provides a unique approach to characterize development based on differentiated cells. This strategy complements those based on the direct analysis of progenitor cells (e.g., single-cell RNA sequencing)⁴¹ in multiple ways. Firstly, it provides information about the long-term fate of progenitors. Secondly, it enables analyzing progenitor states with respect to specific subsets of their progeny which may be of interest, for example due to relevance to a specific disorder. Finally, it can be applied to non-model species wherein accessing somatic mutations from cadavers is more practical than obtaining embryos.

Our results show that only when a progenitor state is sufficiently sampled can its potency and dynamics be meaningfully estimated in one individual and improved by combining biological replicates; estimates for severely undersampled progenitor populations are not meaningful irrespective of the number of replicates. In cases modeled here, adequate sampling often required more than 25% of the actual progenitor population to have descendants among sampled terminal cells ($PS_{cov} > 5$). To meet this sampling criterion in practice, the number of descendants that are analyzed should be, at least, in the same

order as the actual progenitor population size. We propose this as a fundamental rule for retrospective lineage analysis. For prohibitively large progenitor populations, this sampling rule may be satisfied by bottlenecking the number of sampled progenitors using non-random sampling of terminal cells based on anatomical position or other criteria.

In summary, we have described quantitative fate mapping as a framework to characterize complex cell fate dynamics during development using retrospective lineage analysis at a later point in time. This framework is based on somatic mutations—synthetic or natural—that accumulate during development. Robust fate mapping requires good representation of each progenitor population among sampled cells as well as the ability to infer time-scaled phylogenetic trees. This framework facilitates the characterization of cell autonomous and non-cell autonomous genetic and environmental effects on development.

Limitations of the study

The parameters of our models were tailored to mouse embryogenesis. Hence, certain thresholds and cutoffs associated with specific conclusions may be different in later stages of development, other species, and other developmental systems such as organoids. In such cases, the simulation strategies described here can be employed to obtain system-specific values. Our models assume lineage-independent rates of mutagenesis. While this assumption is supported as a first-order approximation^{5,42}, we expect our strategy to be robust to small lineage biases in mutation rates. We have not performed experimental validation of this framework in a model organism, which may reveal factors that were not considered here and require adjustments to ICE-FASE and Phylotime strategies.

STAR Methods

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Reza Kalhor (kalhor@jhu.edu).

Materials availability—The human induced pluripotent stem cell line generated in this study will be shared upon request.

Data and code availability

- FASTQ files from sequencing single-cell hgRNA arrays have been deposited at SRA and are publicly available. Accession numbers are listed in the key resources table. All quantitative fate maps, simulated datasets, inDelphi predictions, and Phylotime-reconstructed phylogenies have been deposited on Zenodo. DOIs are listed in the key resources table.
- R package for QFM and code to reproduce the results is publicly available at <https://github.com/Kalhor-Lab/QFM/> as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture—Human iPSCs were cultured in mTeSR Plus media (STEMCELL Technologies) on plates coated with Matrigel Growth Factor Reduced Basement Membrane Matrix (Corning). Cells were maintained at 37°C and 10% CO₂/5% O₂ conditions with daily media changes. When up to 80% confluent, cells were passaged by dissociation with Accutase (STEMCELL Technologies) and seeded in mTeSR Plus media supplemented with 5 μM blebbistatin (Millipore Sigma).

Lineage-tracing human induced pluripotent stem cell line—We established a clonal iPSC line with 32 hgRNA barcoding sites distributed in its genome as a non-tandem array. The line also includes doxycycline inducible Cas9 to activate barcoding. The EP1 human induced pluripotent stem cell (hiPSC) line was obtained from Bhise et al.⁵², originating from fetal lung fibroblasts obtained from a female donor. A clonal lineage-tracing inducible Cas9 cell line was generated by first transfecting the EP1 cells with plasmids to target and stably insert both a reverse tetracycline-controlled transactivator (rtTA) construct and a tetracycline-dependent Cas9 construct into each of the two copies of the AAVS1 safe harbor locus. The cells were then transfected with a hgRNA PiggyBac library and screened for high numbers of insertions. A KaryoStat™ Assay of the final engineered cell line confirmed that the sample originated from a female and had no chromosomal aberrations when comparing against the reference dataset. Cells were determined to be free of mycoplasma contamination based on MycoAlert® PLUS Mycoplasma Detection Kit (Lonza) test results.

METHOD DETAILS

Definition of quantitative fate map—A quantitative fate map (QFM) is characterized by a topology in the form of a bifurcating or multifurcating time-scaled tree. Each node in the tree is associated with a time, with the root node at time 0 and the terminal nodes at time of sampling. Each edge in the tree represents a progenitor state or terminal type, the starting and end time of the edge represent the duration for which the state existed. If an edge ends at the sampling time, it is a terminal type, otherwise a progenitor state. For progenitor states, the end node of the edge represents a commitment event. For terminal types, the end node of the edge represents the time at which the process is stopped for sample collection. For a bifurcating QFM with Z total terminal types, we denote the progenitor states as P_i , $i = 1, 2, \dots, (Z - 1)$ and the terminal types as T_j , $j = 1, 2, \dots, Z$. Each progenitor state or terminal type has a set of parameters associated with it. We take a progenitor P_i as an example here. Along the edge representing P_i , cells undergo cell division and cell death at rates specific to the state. The doubling time for P_i is denoted d_{P_i} . When not committing, cells can be (i) dividing (ii) not dividing, or (iii) dying with probabilities $\phi_{P_i,D}$, $\phi_{P_i,N}$ or $\phi_{P_i,L}$ respectively. These probabilities are specific to the progenitor state. At the commitment time for P_i , denoted t_{P_i} , the cells commit to the downstream states. For a bifurcating commitment event, when cells of the P_i state commit to the downstream states, say P_j or P_k , it can do so via either via symmetric cell division or asymmetric cell division. For commitment via symmetric cell division, each cell of the P_i state gives rise to either two P_j cells or two P_k cells. For commitment via asymmetric cell division, each cell gives rise to two cells of different downstream states, that is, one cell P_j cell and one P_k cell. The corresponding

commitment mode probabilities are denoted ($p_{P_j P_j}$, $p_{P_k P_k}$, $p_{P_j P_k}$), where $p_{P_j P_j}$ and $p_{P_k P_k}$ are the probabilities of the symmetric modes that commits to P_j or P_k respectively and $p_{P_j P_k}$ is the probability of the asymmetric mode. The commitment bias is defined as the proportion of cells of each of the downstream states that are produced. In the above example, the commitment bias, denoted as b_{P_j} is $(2p_{P_j P_j} + p_{P_j P_k}, 2p_{P_k P_k} + p_{P_j P_k})$. In most fate maps simulated here, $p_{P_j P_k} = 0$ as they did not include a specific asymmetric division mechanism. For the purpose of evaluation, for each bifurcating progenitor state, the proportion that commits to one of the two downstream states is chosen as a single value representing the progenitor state bias and is kept track of consistently throughout. The progenitor population size N_{P_j} is defined as the number of cells at the commitment time.

Definition of time-scaled phylogeny—A time-scaled phylogeny is defined as a rooted, ultrametric, bifurcating phylogenetic tree where branch lengths are in the unit of time and represent the time in between cell divisions. Terminal nodes, or tips, of the tree represent observed cells. Internal nodes in the time-scaled phylogeny represent cell divisions of unobserved progenitor cells. The root node represents the most recent common ancestor (MRCA) of all terminal cells. The length of the root edge is the time until the cell division of the root MRCA. Cophenetic distance is defined for each pair of terminal cells, which is the distance between the cells along the phylogenetic tree. The depth of a node in the phylogenetic tree is defined as the distance of a node to the root plus the length of the root edge. The ultrametric property requires that all tips are equidistant from the root, that is, have the same depth. The total time of a time-scaled phylogeny is defined as the depth of its tips.

Constructing a panel of quantitative fate maps—Generating each fate map in our panel involved several interrelated and iterative steps. These steps were designed to ensure that the fate maps in the panel explore a wide range of possibilities in terms of topology and progenitor state parameters while making sure: (i) the fate map generated resembles early mouse development (ii) the commitment events are placed within our prespecified time window ([2.4,10.9] days) (iii) two consecutive commitment events are at least one doubling time apart. For clarity, these steps are briefly listed here and detailed separately below. First, a candidate topology in the form of a bifurcating tree was generated from one of five categories of varying imbalance. Edges in the bifurcating tree represent progenitor states and terminal types; nodes in the tree represent their commitment events. Second, an ordering of the commitment events was generated. Third, doubling times of progenitor states were drawn, leading to a candidate fate map. Fourth, the timing of commitment events in the candidate fate map was checked against the minimum inter-commitment time and total span criteria. If the candidate fate map met these criteria, it was accepted and its commitment biases and death rates were drawn. Otherwise, a perturbation to the candidate topology was suggested, and steps one through four were repeated until a valid fate map emerged.

Generating candidate topologies: To generate a panel of fate map topologies with varying levels of imbalance, tentative tree topologies were generated from five different categories: (i) ‘perfectly balanced’ (ii) ‘balanced-TBR’ (iii) ‘random’ (iv) ‘pectinate-TBR’ (v) ‘pectinate’. In the perfectly balanced topology, each bifurcation splits all terminal types

into two equal halves. The balanced-TBR topologies are generated by applying one random TBR move to the perfectly balanced topology. The random topologies are generated by creating a sequence of bifurcations, each randomly splitting the set of terminal types into two sets; the 'rtree' function in the 'ape' R package⁵³ was used to generate the random topologies. The random category tends to have topologies of medium balance. The pectinate topology is in the shape of a comb. In pectinate topology, each bifurcation always splits the terminal types such that one of the splits has exactly one terminal type (Figure 1C, **bottom left**). The pectinate topology is the maximally unbalanced topology for a fixed number of terminal types. The pectinate-TBR topology is generated by applying a number of TBR moves to the pectinate topology.

Generating event ordering: In the second step, we generated an ordering of commitment events following a previously published method⁴⁴. Briefly, to get an ordering of events in the entire topology, a 'shuffling' at each bifurcation was generated. A shuffling orders the events between two subtrees of a bifurcation without ordering within each subtree. For example, two shufflings of a given event are illustrated in Figure S8A. Events from the left subtree are shown as hollow circles and events from the right subtrees are shown as solid ones. The sequence of solid and hollow circles is defined as a shuffling at the bifurcation. A set of shufflings for every bifurcation (internal node) of a tree determines the exact ordering of its commitment events.

Generating doubling times and exact commitment times: Given the event ordering, doubling times for each state/type are drawn. To have overall agreement with the mouse development but also allow variabilities across cell states/types, the following scheme was used. The root state has a doubling time of 0.6 days. All other progenitor states have the doubling time drawn from a uniform distribution whose center is determined by the time ranking of its commitment event. Specifically, for an event ranked i -th out of total $I - 1$ commitment events, the doubling time follows the uniform distribution below:

$$d_{(i)} \sim \text{uniform}\left(0.5 - (i - 1) * \frac{0.55 - 0.4}{I - 2}, 0.6 - (i - 1) * \frac{0.55 - 0.4}{I - 2}\right)$$

For example, for a 16-terminal type fate map, the earliest commitment non-root progenitor state has doubling time sampled from uniform(0.49, 0.59) and the latest committing progenitor state has doubling time sampled from uniform(0.35, 0.45). For all terminal types, the doubling time is sampled from uniform(0.35, 0.45) (Figures S1F,G). These doubling rates generally agree with rates reported in the literature²³.

Next, we generated the duration between consecutive events in the commitment event ordering. In our model, consecutive commitment events were required to be at least one cell division (doubling time of the earlier state) apart. In addition, all of the commitment events need to fit in a predefined time window $[l_0, l_1]$. These requirements place constraints on the duration in between commitment events. Such a constraint can be formalized as follows:

Suppose there are I total events, where the duration between the $(i-1)$ -th and i -th event in the event ordering is y_i . Let the doubling time of progenitor state i be d_i , then for an edge in the fate map topology that connects the earlier event j to the later event k , we have

$$\sum_{i=j+1}^k y_i - d_j \geq 0 \text{ and } y_i \geq 0$$

There are $I-1$ total such constraints, one for each edge that is not the root edge. Next, we used linear programming to find the vector of y_i 's such that $\sum_i y_i$ is minimized, the minimum value of $\sum_i y_i$ is denoted y_{total} . When generating the event timing, we tried to place the events within the given time window: $[l_0, l_1]$ with its total length being $l_{\text{total}} = l_1 - l_0$. For a valid placement of events to exist, one needs to have $y_{\text{total}} \leq l_{\text{total}}$. At this stage, we checked if this condition is met. If it was not, the current value of y_{total} was recorded and the candidate fate map topology is perturbed by applying one tree bisection and reconnection (TBR) move. TBR is a tree rearrangement that detaches a subtree from the main tree at an interior node and then reconnects to another random branch. After the TBR move, all the above steps were repeated and a new y_{total} values was computed, denoted y_{total}' . If $y_{\text{total}}' < y_{\text{total}}$, then the newly perturbed topology was accepted as the tentative fate map topology. Otherwise, the newly generated perturbation was discarded, and another perturbation was generated. The process was repeated until a $y_{\text{total}} \leq l_{\text{total}}$ was found. With the new l_{total} found, we tried to place events uniformly within the time window by distributing the remaining duration $z_{\text{total}} = l_{\text{total}} - y_{\text{total}}$ to each interval evenly, which was achieved by sampling from the following Dirichlet distribution:

$$\left(\frac{z_1}{z_{\text{total}}}, \frac{z_2}{z_{\text{total}}}, \dots, \frac{z_{I-1}}{z_{\text{total}}} \right) \sim \text{Dirichlet}(5, 5, \dots, 5)$$

Finally, the commitment time for the u -th event is $l_0 + \sum_{i=1}^u (y_i + z_i)$ (Figure S1D).

Generating commitment bias and death rate: The raw commitment biases were drawn from Beta(5, 5), the raw values were later adjusted to actual values based on how many cells were split into each of the downstream fates. For example, when the progenitor population had 13 cells, and the raw commitment bias was (0.63, 0.37) based on which $[13 \times 0.63] = 8$ cells are committed to one downstream fate versus $[13 \times 0.37] = 5$ cells to the other, with $[\cdot]$ being the rounding operation that rounds to the nearest integer. Then the actual commitment bias was adjusted to (8/13, 5/13). We required that there are at least 4 cells committing to each downstream fate, this requirement resulted in the commitment bias to be closer to 0.5 for smaller progenitor populations (Figure S1E). The cell death rates were drawn from Uniform(0.02, 0.08).

Generating the panel of fate maps: The balance of a fate map topology can be measured by the Colless index. The Colless index of a bifurcating tree is defined as the sum of the balance values of its internal nodes, where the balance value of an internal node is defined as the absolute value of the difference between the number of descendant tips of its pair of daughter nodes²². To get a representative panel of fate maps topologies in terms of

balance, we generated 1,000 fate maps based on topologies from the balanced-TBR, random and pectinate-TBR categories with either 16, 32 or 64 terminal types. After computing the Colless index for each topology, we classified the generated topologies into bins of different Colless index values by increments of 20. Next, within each bin, we randomly selected five (or fewer in bins with fewer than five in total) fate maps (Figure S1C). Finally, we add to this list the perfectly balanced topology and the pectinate topology. Note that because of the constraints of the minimum duration in between consecutive commitment events and the doubling times specified, the pectinate topology does not allow a valid placement of commitment events within the [2.4, 10.9] time window for the fate maps with 32 and 64 terminal types, so it was only added in the case of 16 terminal types. The above procedure produced 53 fate maps with 16 terminal types, 108 fate maps with 32 terminal types and 170 fate maps with 64 terminal types. To see if the steps used here generated topologies that show good variabilities, we embedded the tree topologies into the 2-dimensional space by applying multi-dimensional scaling to the pairwise KC_0 distances. In the embedding, we observed that the first principle coordinate correlated well with the Colless index, and in addition, the second principle coordinate also had good variations, indicating that our topologies cover a range of possibilities (Figures S1A,B).

Generating a count graph for QFM—To carry out computation and generate time-scaled phylogenies from a QFM, an abstract count graph was constructed based on the QFM's specifications, which is a detailed representation of the computations involved (Figure S2).

Each node in the count graph contains a group of cells. To distinguish the terms, we call a node in the count graph a 'count node'. In addition to containing cells, each count node has other relevant attributes. It has a time window, which specifies the time during which its cells exist, it also has a state/type assignment, which specifies which progenitor state or terminal type its cells belong to. Generating a count graph for a QFM starts at its root ($t = 0$) with a single count node. The count nodes in the next time window are generated by applying four operations (which are indicated by **bold** font in text below) to the starting count node and its cells according to the QFM specifications to generate the count node(s) at the next time window. This process is then repeated on the nascent count nodes one round after the other until the entire count graph is created. The detailed implementation of each operation is given in the later section (Figure S2, Step 1).

During each time window, which is the length of a cell cycle, based on if the commitment time of the progenitor state has been reached in the QFM, each count node undergoes one of two processes before creating count node(s) in the next time window:

1. a proliferate-only process
2. a commit and proliferate process

Initially, the count node is in a "default" mode (labeled T), representing its total population of cells during the time window. If undergoing process 1 (proliferate-only), the count node is split into three sub-count nodes that represent cells that are (i) doubling (labeled D) (ii) not doubling (labeled N) or (iii) dying (labeled L). Specifically, cells (of cell state P_i) with

N total cells are **split** into three sub-count nodes of different proliferation modes, according to probabilities c_{D,P_i} , c_{N,P_i} and c_{L,P_i} . Here, the **split** operation splits a count node into three sub-count nodes:

$$(N_D, N_N, N_L) = \text{split}(N, (c_{D,P_i}, c_{N,P_i}, c_{L,P_i}), m_D = m_N = m_L = 0), c_{D,P_i} + c_{N,P_i} + c_{L,P_i} = 1$$

where N_D , N_N , N_L are the counts of the sub-count nodes, c_{D,P_i} , c_{N,P_i} , c_{L,P_i} are the doubling, non-doubling and dying probabilities, and m_D , m_N and m_L are the minimum number of cells in each sub-count node after the split.

Then, the doubling sub-count node **doubles** in number and is **merged** with the non-doubling sub-count node to give rise to the default count node of the next time window, which starts one doubling time c_{P_i} later than the current time window. The dying sub-count node is lost in the process. The **double** operation doubles the count, and the **merge** operation adds up the counts of sub-count nodes.

If undergoing process 2 (commit and proliferate), cells in the default count node are first **split** into one of the commitment modes and subsequently **double** or **asymmetric double** depending on which commitment mode they are assigned to in the QFM. The **asymmetric double** operation originates from a sub-count node and gives rise to two sub-count nodes of the same size but are of two different downstream states/types. Suppose cells (of some progenitor state P_i) are committing during a time window to two downstream fates P_j and P_k , then the count node in default mode gets split into three sub-count nodes of different commitment modes. There are three different commitment modes when commitment is a bifurcation: two symmetric commitment modes $P_i-P_jP_j$ and $P_i-P_kP_k$ (one each for the two downstream fates) and one asymmetric commitment mode $P_i-P_jP_k$. Cells in one of the two symmetric commitment modes each gives rise to two cells of the same downstream types. Cells in the asymmetric commitment mode each gives rise to two cells of two different downstream types. Let the probabilities of the two symmetric commitment modes be $p_{P_jP_j}$, $p_{P_kP_k}$ and the probabilities of the asymmetric commitment mode be $p_{P_jP_k}$. Specifically, during a time window when cells are in process 2 (commit and proliferate), when there are N cells in the default count node, the cells are split into the commitment sub-count nodes via the **split** operations:

$$(N_{P_jP_j}, N_{P_kP_k}, N_{P_jP_k}) = \text{split}(N, (p_{P_jP_j}, p_{P_kP_k}, p_{P_jP_k}), (m_{j,j}, m_{j,k}, m_{k,k})) = (4, 0, 4)$$

Where $N_{P_jP_j}$, $N_{P_kP_k}$, $N_{P_jP_k}$ are the counts of the sub-count nodes, $p_{P_jP_j}$, $p_{P_kP_k}$, $p_{P_jP_k}$ are the probabilities for the commitment modes, and $m_{P_jP_j}$, $m_{P_jP_k}$, $m_{P_kP_k}$ are the minimum number of cells in each sub-count node after the split. To ensure that there's no progenitor population of very small size, we require that there be at least four cells in the symmetric modes in the main panel.

The commitment bias is defined as $(2p_{P_jP_j} + p_{P_jP_k}, 2p_{P_kP_k} + p_{P_jP_k})$, the two proportions sum to one and are interchangeable, we refer to one of the two sides as the commitment bias p_{P_i} of the P_i state, which is kept consistent throughout. In the main panel of experiments,

all probabilities of the asymmetric commitment modes are set to zero. Subsequently, cells of the symmetric commitment modes **double** and then **merge** with the daughter cells of the asymmetric commitment modes for each of the downstream states. For P_j , $N'_{P_j} = 2N_{P_jP_j} + N_{P_jP_k}$ and for similarly for P_k , $N'_{P_k} = 2N_{P_kP_k} + N_{P_jP_k}$.

Backpropagation to generate progenitor population sample sizes—To generate a phylogeny for a set of sampled cells based on the quantitative fate map, the number of sampled cells was generated at each (sub-)count node in the count graph (Figure S2, Step 2). This process draws, for each (sub-)count node in the count graph, an additional quantity called its sample size. A progenitor cell is sampled if any of its progeny at the sampling time is sampled. The total number of sampled cells is defined as its sample size. The process was accomplished by propagating backwards in the count graph. During the backward propagation, each forward operation is replaced with a corresponding backward reverse operation. The detailed implementation of each operation is given in the later section, and the operations are again indicated with bold text. The sample size for the terminal population is set to the sample size of the experiment. For fixed sampling, the size is the same for all the terminal types. For proportional sampling, the total sample size is distributed to each terminal type according to their abundance, which is done using the **stochastic split** operation. The **stochastic split** operation is the reverse operation of **merge**. Given the sample size of the merged population, the reverse operation draws the number of cells sampled in each mode based on a multivariate hypergeometric distribution. In addition, the reverse operation for **split** is **merge**, which adds up the sample size of count nodes of different modes to the sample size of a count node with the default mode. The reverse operation for **double** is **stochastic coalesce**: given the number of sampled cells in a doubled population, the operation draws the number of cells that are sampled in the pre-doubling population. The reverse operation for **asymmetric double** is **stochastic asymmetric coalesce**: given the number of cells sampled in each of the populations after asymmetric doubling, the operation draws the number of cells sampled in the population pre-asymmetric double. In summary, the backpropagation process generates the sample size for each (sub-) count node in the count graph given the sample size of the terminal populations.

Generating phylogeny of sampled cells—Given sample size of each count node derived in the previous section, we moved to generate the phylogeny for a group of sampled cells. Before the individual cells were generated and connected to construct a cell phylogeny, the count graph was first reorganized (Figure S2, Step 3a). At this step, we are only concerned with the sample sizes of the count nodes, not the population sizes. During the reorganization, the three sub-count nodes during process 1 (proliferate only) were replaced with two new sub-count nodes, and the three sub-count nodes during process 2 (commit and proliferate) were replaced with five new sub-count nodes. The reorganization is done based on if a sampled cell in the current sub-count node either gives rise to one or two cells in the next (sub-)count node. Note that the sample size(s) for the next count node(s) determines exactly how many cells in the current count node give rise to one and how many give rise to two cells. For example, when there are 5 sampled cells in the current count node of the doubling sub-count node and there are 8 cells in the doubled sub-count

node, then exactly 3 cells in the current count node give rise to two cells at the doubled sub-count node and 2 cells give rise to one. As another example, if the $P_i P_j P_k$ sub-count node have 5 cells sampled and the P_j doubled sub-count node has 3 cells sampled and the P_j doubled sub-count node has 4 cells sampled, then there are exactly 2 cells with both daughters sampled, 1 cell with one A daughter sampled and 2 cells with one B daughter sampled. We call progenitor cells with both daughters sampled ‘coalesced’, and progenitor cells with only one daughter sampled non-coalesced. Subsequently, the sub-count nodes are reorganized based on if cells are coalesced or non-coalesced as well as cells of which cell or state they give rise to.

For process 1, coalesced cells in the doubling sub-count node are organized into a new ‘coalesced’ sub-count node (labeled C) and non-coalesced cells in the doubling sub-count node sampled are combined with cells in the non-doubling sub-count node and organized into a new non-coalesced (labeled NC) sub-count node. Cells in the death sub-count node are lost. The original doubled sub-count node is replaced with the coalescing (labeled CI) sub-count node (Figure S2, Step 3a, top example).

For process 2, coalesced cells in each of the three original sub-count nodes ($P_i P_j P_k$, $P_i P_k P_k$ and $P_i P_j P_k$) that give rise to two cells in the next count nodes are organized into their respective new sub-count nodes (labeled the same name). Two new sub-count nodes were added that represent cells that give rise to only one cell (of either of the downstream types) in the next count node, (labeled $P_i P_j$ and $P_i P_k$). Non-coalescent cells in the $P_i P_j P_k$ sub-count node are merged with non-coalesced cells that give rise to only one P_j cell in the $P_i P_j P_k$ sub-count node. Similarly, non-coalesced cells in the $P_i P_k P_k$ sub-count node are merged with cells that give rise to only one P_k cell in the $P_i P_j P_k$ sub-count node.

The reorganized count graph (Figure S2, Step 4a, right) specifies the number of sampled cells at each stage but does not distinguish between individual cells. To generate the sampled cell phylogeny, we next listed individual cells and specified which cells give rise to one or two sampled cells at the next time point without specifying which cells at the next stage they were connected to (Figure S2, Step 3b). Next, cells in an earlier time window were connected to cells in the later time window to make up the phylogeny (Figure S2, Step 4a). Finally, the phylogeny was further simplified by removing internal nodes that give rise to one cell only (Figure S2, Step 4b).

Definition of commitment time and progenitor population size—In our model, the fate commitment modes are assigned at the beginning of the process 2 time (commit and proliferate) window, before cells proliferate to become the downstream states or types. In fact, the fate decision could have been made any time during the previous cell cycle. Hence, the exact commitment time could be chosen as either at the beginning of the previous time window or at the end of the window. Similarly, the progenitor state population size may be defined as the number of cells at either time point. As detailed in Figure S8B, it can be observed that, for symmetric commitments, the time of fate separation (where daughter cells are exclusively one of the downstream states) on the phylogeny is, at the latest, at the beginning of the cell cycle (t_1). On the contrary, for commitment by asymmetric cell division, fate separation is at the end of the cell cycle (t_2). Therefore, we define

the beginning of the window as the commitment time for progenitor states undergoing symmetric mode of commitment, and the end of the window as the asymmetric commitment time. The progenitor population sizes for symmetric and asymmetric commitment modes are defined as the number of cells N_1 and N_2 at the corresponding commitment times.

Definition of operations to generate count graph—The operations used in the count graph generation and backward propagation of sample sizes are formally defined here. Four deterministic operations are defined for generating the population counts in quantitative fate maps.

split: $(N, (p_1, p_2, \dots, p_I), (m_1, m_2, \dots, m_I)) \mapsto (N_1, N_2, \dots, N_I)$

The **split** operation splits some total count N into a vector of output counts N_1, N_2, \dots, N_I based on the input probabilities p_1, p_2, \dots, p_I and ensures the post-split counts: (i) add up to the total count, (ii) are larger than some minimum values m_1, m_2, \dots, m_I and (iii) are all integers. First, let m be the smallest m such that $p_m \geq p_i$ for $i \in \{1, 2, \dots, I\}$. Then, the $N_i = \max([Np_i], m_i)$ for all $i \leq m$, and $N_m = N - \sum_{i=1}^{m-1} N_i$, where $[x]$ is the rounding operation that returns the integer that is closest to x .

merge: $(N_1, N_2, \dots, N_I) \mapsto N$

The **merge** operation takes a vector of input counts and them, i.e. $N = \sum_{i=1}^I N_i$

double: $N \mapsto 2N$

The **double** operation multiplies the input count by two.

asymm double: $N \mapsto (N, N)$

The **asymmetric double** operation takes the input count and produces two input counts of the same size.

For each of the four forward operations, a reverse operation is defined for generating the sample sizes based on the population sizes.

stochastic split: $(S, (N_1, N_2, \dots, N_I)) \mapsto (S_1, S_2, \dots, S_I)$

stochastic split is the reverse operation of **merge**. (S_1, S_2, \dots, S_I) follows a multivariate hypergeometric distribution with the following probability mass function

$$p(S_i = k_i | N_i, S) = \frac{\prod_{i=1}^I \binom{N_i}{k_i}}{\binom{\sum_i N_i}{S}}$$

merge is the reverse operation of **split**.

stochastic coal: $(S, N) \mapsto S'$

stochastic coalesce is the reverse operation of **double**. When S cells are sampled from $2N$ cells that are direct descendants of N cells, S' is the number of cells sampled in N cells. Here, the number of coalescences $C = S - S'$, follows the distribution with the below probability mass function:

$$p(C = z | S = k) = \binom{n}{z} \binom{n-z}{k-2z} 2^{(k-2z)} / \binom{2n}{k}$$

After drawing C , S' is computed with $S' = S - C$.

stochastic asymm coal: $((S_1, S_2), N) \mapsto S'$

stochastic asymmetric coalesce is the reverse operation of **asymmetric double**. When N parent cells double, and each cell gives rise to two cells of two different downstream states/types, and S_1 out of N cells of type A and S_2 of N cells of type B are sampled. Then the number of cells sampled in the parent population is S' . Here, the number of coalescences $C = S_1 + S_2 - S'$ follows the distribution with the below probability mass function:

$$p(C = z | S_1 = k_1, S_2 = k_2) = \binom{n}{z} \binom{n-z}{k_1-z} \binom{n-k_1}{k_2-z} / \binom{n}{k_1} \binom{n}{k_2}$$

After drawing C , S' is computed with $S' = S_1 + S_2 - C$.

Fate map topology reconstruction with FASE—To determine if a node is a FASE (SEparation (FASE)), a list of unique terminal types is generated for all internal nodes of the phylogeny which constitutes its observed fate. A node is classified as a FASE (for at least one pair of terminal types) if either of its daughter nodes has an observed fate that is less potent. FASEs are identified across the entire phylogenetic tree. Next, for each FASE, all pairs of terminal types that the FASE separates are listed. A pair of terminal types is separated if one terminal type is seen in only one branch of the node and the other type only in the other. Then, for a pair of terminal types, the mean depth of the FASEs that separated the terminal fates were computed, referred to as the FASE time. If no FASE existed for a pair of terminal types, the FASE time was taken to be zero. Finally, the FASE distance between a pair of terminal types is equal to the difference between total time of the phylogenetic tree and their average FASE time. To reconstruct the topology from the full distance matrix, the *upgma* function from the “Phangorn” package⁵⁴ was applied, which is a wrapper of *hclust* in base R.

Fate map topology reconstruction with SPS—Computation of shared progenitor scores was implemented as specified elsewhere⁶. The shared progenitor score or SPS was calculated between two terminal types as the number of internal nodes in the phylogeny that have both terminal types scaled by the number of total terminal types each internal node contributes to. The score was computed by going through all the internal nodes in the phylogeny. For each internal node, let T_1, T_2, \dots, T_L be the list of terminal types it gives rise to, with L being the length of the list, then its contribution to all pairs of terminal types derived from the list is $1/(2^{L-1})$. For example, if an internal node is capable of $\{T_1, T_2, T_3\}$,

and let $SPS_{[T_i, T_j]}$ denote the shared progenitor score for the pair of terminal types T_i and T_j ; then for we can compute the contribution of the internal node as $1/(2^{3-1}) = 1/4$. Then

$$SPS_{[T_1, T_2]} + = (1/4)$$

$$SPS_{[T_1, T_3]} + = (1/4)$$

$$SPS_{[T_2, T_3]} + = (1/4)$$

After iterating over all internal nodes, a matrix of shared progenitor scores M_{SPS} was generated. The SPS matrix was converted to a similarity matrix by $1 - M_{SPS}/\max(M_{SPS})$. Finally, UPGMA clustering was applied to the similarity matrix to obtain the fate map topology.

Random topology generation—For providing a baseline of topology reconstruction errors, we generated a random topology for each experiment compared in Figures 2F and 5A using the same method as described in the fate map reconstruction section. Briefly, the ‘rtree’ function in the ‘ape’ R package⁵³ was used. This method creates bifurcations by randomly assigning terminal types to each side of the bifurcation based on draws from a uniform distribution.

Mapping of inferred progenitor states to true progenitor states—Upon initial reconstruction of fate map topology, the correspondence between the inferred progenitor states and the true progenitor states in the original fate map is not known. For downstream evaluations, true and inferred progenitor states were matched based on their potency and commitment patterns. An inferred progenitor state was considered correctly resolved if there existed a true progenitor state that met the following conditions:

- i. the true progenitor state had the exact same potency as the inferred progenitor state.
- ii. the immediate downstream states or types of the true progenitor state had the exact same potencies as the immediate downstream states/types of the inferred progenitor state.

Node state assignment in time-scaled phylogeny—One characteristic of a progenitor state is its potency: the set of terminal types it can lead to. Each node in the time-scaled phylogeny also had an observed potency determined by the set of states its progeny covers. The inferred progenitor state of a node of the phylogeny was assigned based on its potency: the node was assigned an inferred state that had the same potency as itself. If no such state existed in the fate map, then it was assigned the least potent state in the fate map that was more potent than the node but included all observed fates of the node in its potency.

Commitment time inference with ICEs—To infer commitment time of a progenitor state, a set of Inferred Commitment Events (ICEs) were identified. A node in the time-scaled phylogeny was considered an ICE if both daughters had a different assigned state than itself. Unlike FASEs, which are defined for each pair of terminal fates, ICEs are defined with respect to inferred progenitor states. If a node is an ICE, it is also a FASE for at least some pair of terminal states, but the opposite is not necessarily true. Each ICE was associated with a progenitor state. The depths of all ICEs associated with a state defined the ICE times. The mean of ICE times was used as an estimate for the commitment time. In the case where inferred commitment times of the downstream state was earlier than that of the upstream state, the commitment time of the downstream state was set to that of the upstream state. This situation is indicative of the lack of evidence supporting the relative ordering the two commitments in question.

Progenitor population size and commitment bias inference—To identify the population present at the commitment time of a progenitor state, first, a set of extended states was defined for each inferred progenitor state. The extended states included the state itself, its upstream states up to root and its downstream states down to the terminal types. Next, a state path was constructed on fate map topology for each branch that spanned the commitment time of the progenitor state: the state path included all the inferred progenitor states in the fate map topology between the state of the incoming node (closer to the root of the phylogeny) and the state of the outgoing node (closer to tips of the phylogeny) of the branch (Figure 3G). A branch was considered associated with the progenitor state if its state path was a subset of the extended states of the progenitor state. To estimate the progenitor population size, the collection of incoming nodes of the associated branches were counted (Figure 3H).

To quantify the bias of a progenitor state's commitment, each associated branch was further classified as (i) committing to one of the downstream states, or (ii) uncommitted. The classification was made based on the state path; if the state path covered one of the immediate downstream states, it was classified accordingly. Otherwise, it was classified as uncommitted. For commitment bias, the proportion of committed branches committing to each downstream state of the inferred progenitor state was used.

Barcoding mutagenesis model—Let there be $i = 1, 2, \dots, I_{\text{total}}$ barcoding sites. A barcoding site is unmutated at the beginning (i.e., $t = 0$) and once activated, starts to accumulate heritable mutations over time. Mutation events happen independently for each site according to a Poisson point process with a constant rate after activation (Figure 4A). Let the mutation rate of the Poisson process be λ_i for site i , then the probability of having a mutation event in a small time window $[t, t + \Delta t]$ is $\lambda_i \Delta t$. In our model, once a site has mutated, it can no longer mutate further. A mutant allele is inherited from a cell by all its descendants. The Poisson point process has the memoryless property, which means that if no mutation event has happened up until time t , then the probability of getting a mutation event in the time window $[t, t + \Delta t]$ is $1 - e^{-\lambda_i \Delta t}$, which does not depend on the value of t .

When a mutation event occurs, an unmutated active copy of the site is converted into one of many possible mutated inactive alleles, each with a distinct emergence probability (Figure

4A). Let $A_{i1}, A_{i2}, \dots, A_{iJ_i}$ be the set of possible alleles for site i , and $a_{i1}, a_{i2}, \dots, a_{iJ_i}$ be their corresponding probabilities, then a mutant allele X_i of site i created by a mutation event has $P(X_i = A_{ij}) = a_{ij}$. To summarize, the parameters of the barcode mutagenesis model include the mutation rates $\{\lambda_j\}_{j=1, \dots, J}$ and each site's mutant allele emergence probabilities $\{a_{i1}, \dots, a_{iJ_i}\}_{i=1, \dots, I}$.

Parametrization of barcode mutagenesis model with MARC1 mouse data—

To simulate lineage barcodes that mimic a realistic system, the mutagenesis model was parametrized based on the MARC1 (Mouse for Actively Recording Cells 1) system⁵ wherein extensive embryonic barcoding data are available³⁰. In MARC1 mice, somatic mutations are induced in tens of independent homing guide RNA loci (hgRNAs)²⁹. We estimated the mutation rates of MARC1 hgRNAs (i.e., rate of the Poisson process) using embryonic time course data (Figure S4A). We estimated emergence probabilities of mutant alleles for each hgRNA by adapting the inDelphi algorithm that predicts CRISPR-Cas9 mutations³¹. The details of this estimation can be found in the Quantification and Statistical Analysis section. We compared and verified inDelphi's predictions against published MARC1 data (Figure S4B). To test how well the results resembled actual lineage barcoding data, we simulated barcoding in whole-mouse embryos for E3.5 to E16.5 in samples of 2,000 cells (or fewer when there were fewer than 2,000 cells in the organism) and compared the results to that of experiments. See Figures S4A–G and their legends for the results and their interpretation.

Parameters for simulating MARC1 lineage barcodes—Parameter estimates from the MARC1 data were used to simulate lineage barcodes from time-scaled phylogeny. The corresponding estimated mutation rates and inDelphi predicted allele emergence probabilities for each MARC1 hgRNA were used as parameters for the mutagenesis model. For reconstructing Phylotime phylogeny in Figure 5A,B, 50 hgRNAs of the 'intermediate' or 'fast' category were randomly sampled from the MARC1 pool of hgRNAs for each simulation.

Phylotime for reconstructing time-scaled phylogeny from lineage barcodes—

Our approach to reconstructing time-scaled phylogenetic trees for thousands of cells was based on a maximum likelihood estimation of pairwise temporal distances between cells. Given a pair of terminal cells, with their MRCA being N_1 , the branch length parameter t_{N_1} was estimated, which is the time since the MRCA of the two cells (Figure S4H).

For a barcoding site i with a mutation rate of λ_i and probabilities $a_i = (a_1, a_2, \dots, a_{J_i})_i$ of mutating into alleles $(A_1, A_2, \dots, A_{J_i})_i$, the likelihood of observing the given alleles M_{iC_1} and M_{iC_2} in a single barcoding site in two cells (c_1, c_2) is the sum of two terms:

$$p_i = p_{i, N_1, 1} + p_{i, N_1, 0}(p_{i, C_1} p_{i, C_2}) \quad (\text{Eq. 1})$$

The first term, $p_{i, N_1, 1}$, is the probability that a mutation has occurred before the MRCA, leading to identical alleles in both cells. The second term is the probability of observing the allele in each terminal cell respectively (p_{i, C_1} and p_{i, C_2}) conditional on no mutation occurring before the MRCA (the probability of which is $p_{i, N_1, 0}$).

For the first term, we first define the term $p_{i, N_1, A_{ij}}$ for each mutant allele A_{ij} , where $p_{i, N_1, A_{ij}} = \left(1 - e^{-\lambda_i t_{N_1}}\right) a_{ij}$ if both C_1 and C_2 carry allele A_{ij} for the site, and $p_{i, N_1, A_{ij}} = 0$ otherwise. Then the first is calculated as:

$$p_{i, N_1, 1} = \sum_{j=1}^{J_i} p_{i, N_1, A_{ij}}$$

which is the sum over all probable mutations, where at most one term can be non-zero. For the second term, we have

$$p_{i, N_1, 0} = e^{-\lambda_i t_{N_1}}$$

and

$$p_{i, c} = \left(1 - e^{-\lambda_i (T - t_{N_1})}\right)^{1 - \text{Ind}(M_{ic} = "0")} \prod_j a_{ij}^{\text{Ind}(M_{ic} = A_{ij})} \left(e^{-\lambda_i (T - t_{N_1})}\right)^{\text{Ind}(M_{ic} = "0")}$$

where M_{ic} denote the allele observed for $c = C_1, C_2$ and "0" denotes an unmutated allele, and $\text{Ind}(\cdot)$ is the indicator function.

Because barcoding sites in our model are assumed to be independent, the likelihood for the set of alleles observed in all barcoding sites is then the product of their individual likelihoods:

$$p(\{\lambda_i, a_i\}_i, t_{N_1}) = \prod_i p_i \quad (\text{Eq. 2})$$

To get estimates of pairwise distance, or equivalently, time since MRCA between two cells (t_{N_1}), we first plugged in estimates of mutation rates and allele emergence probabilities.

In the simulation experiments in this work, the true values of mutation rates and allele emergence probabilities were plugged in. In actual experiments, we suggest using naïve estimates for the mutation rate: $\hat{\lambda}_i = -\frac{1}{T} \log(1 - F_i)$, where F_i is the mutated fraction for site i and T is the total time from the start of the experiment to the sample collection. For mutant allele emergence probabilities, we suggest using estimates obtained from independent experiments or predictions such as inDelphi. If no such information is available, we suggest

using a uniform prior, i.e. $a_j = 1/Jj = 1, \dots, J$. To get the optimal value of t_{N_1} , the following score equation was solved using the Newton-Raphson method:

$$\frac{d \log(p(t_{N_1}))}{dt_{N_1}} = 0 \quad (\text{Eq. 3})$$

The distance between the two cells is $2(T - t_{N_1})$. Once all pairwise distances were computed with the above method, we applied UPGMA hierarchical clustering to derive a phylogenetic tree wherein branch lengths represent actual time. We called this approach Phylotime.

Phylogenetic tree inference with Hamming distance and Cassiopeia—The Hamming distance between barcodes of two cells is defined as the number of sites for which the two cells do not share the same allele. To infer the phylogeny, UPGMA was applied to the pairwise Hamming distance matrix. For computing KC_1 distance, the total length of the inferred tree is scaled to the sampling time. Phylogeny was also inferred with Cassiopeia, a phylogenetic reconstruction software package which was designed for barcode-based reconstruction³⁵. Cassiopeia was run with default parameters with a single thread. The software uses a greedy top heuristic and an integer linear program bottom solver. We allowed Cassiopeia up to 24 hours to solve each tree. The Cassiopeia-reconstructed tree lacked branch lengths and also contained multifurcations that were not fully resolved. The multifurcations were resolved randomly using the ‘multi2di’ function from the ‘ape’ R package⁵³, initial branch lengths were assigned to the tree using the Grafen option, and the ‘chronos’ function was applied to estimate the final branch length.

Simulations of experimental readout errors—To assess how quantitative fate mapping may be affected by readout errors, two common modes of experimental error were considered. These errors include allele dropout, when a fraction of barcoding sites is not detected in every single cell. This form of error is common in experiments where barcoding sites are directly amplified from a single cell’s genome or transcriptome⁸. To simulate allele dropouts, alleles were set to missing completely at random (MCAR), meaning that missing happens with the same probability for all cells and all sites. We also considered allele switching, an error that occurs when an allele from one cell is assigned to another. This form of error can emerge if template switching takes place during barcoding amplification, a possible outcome for synthetic barcoding loci that have a high degree of homology³⁰. To simulate allele switching, we considered a single barcoding site where there are N total cells. For the site, the number of cells with different alleles were first counted: N_1, \dots, N_J cells each with a different unique allele, the counts were then normalized to proportions $f_j = N_j/N$. Given the probability of allele switching p_{switch} the number of cells with error was calculated as $N_{\text{error}} = \text{ceiling}(Np_{\text{switch}})$. To set the error alleles, the N_{error} cells were first drawn, and alleles for those cells were set randomly by a multinomial distribution where the probabilities were the f_1, f_2, \dots, f_J computed earlier. The process was repeated for each barcoding site.

In each case, we applied the error to 5, 10, 20, 30, 40, or 50% of the data from our panel of 3,310 simulated barcoding experiments then applied Phylotime followed by ICE-FASE to reconstruct quantitative fate maps. For allele dropout, we used a strategy to impute missing alleles before applying Phylotime. The strategy, detailed in the next section, predicts missing alleles for each barcoding site using a machine learning algorithm (XGBOOST)⁵⁵ while leveraging information from all the other sites, and it does so sequentially from the site with the least amount of missing data to the site with the most.

Imputation of missing alleles with XGBoost—Given a character matrix that is the cell by site matrix where each element in the matrix is the allele observed, the missing percentages of all barcoding sites were computed first. Next, each site was imputed one by one going from the one missing in the fewest cells to the one missing in the most cells, using information from all other sites (Figure S8C).

Specifically, when imputing missing alleles of a single site, an ‘XGBoost’ model with multinomial ‘softmax’ objective was trained where all the cells whose alleles were observed for the site⁵⁵ were used as the training examples, and all sites that is not the one to be imputed were used as training features, and treated as categorical variables, where each mutant allele is a category. The design matrix of the model was constructed by one-hot encoding each site, treating the missing in the other sites as unmutated, and subsequently column concatenating all one-hot matrices. The parameters ‘max_depth=4’ and ‘nrounds = 20’ were used for the XGBoost model.

Simulation of experiments with different sampling time—A progenitor field may be studied by obtaining its descendant cell types immediately after its development or at a later time point. For instance, development of the primary germ layers (mesoderm, ectoderm, endoderm) during gastrulation may be analyzed by obtaining these cell types immediately after gastrulation (~E7 in mouse) or in late gestation (~E15) by obtaining neuronal cells as ectoderm descendants, muscle cells as mesoderm descendants, and hepatocytes as endoderm descendants. To investigate the effect of sampling time, we simulated phylogeny and barcoding with 50 hgRNAs for all our 16-terminal type fate maps while sampling 100 cells from each terminal type at days 11.5, 12.5, 13.5, 14.5, or 15.5 days (Figure S5E), repeating each condition twice (53 fate maps × 2 sampling strategies × 2 repeats). We then applied Phylotime and ICE-FASE to obtain a quantitative fate map in each case and compared quantitative fate map reconstruction with terminal cell types sampled at various time points after the last commitment (Figures S5F–I). We also applied ICE-FASE to the true phylogeny for comparison (Figures S5J–M). For the results and detailed interpretation see Figures S5E–M and their legends.

Simulation of experiments with increased cell death—Because cell death is a prevalent developmental control mechanism³⁶, we assessed how it affects quantitative fate mapping. In our fate maps with 16 terminal types, we set cell death per division to increasing levels for either all terminal types or all progenitor states with at least 500 cells at their times of commitment (Figure S5N). After simulating barcoding outcomes and applying Phylotime and ICE-FASE, we compared fate map reconstruction accuracy as a function of

cell death rates. For the results and detailed interpretations, see Figures S5N–S and their legends.

Simulation of experiments with commitment via symmetric vs asymmetric divisions—Next, we considered stereotyped cell fate commitment involving asymmetric cell divisions, analogous to those seen in fruit flies eye development³⁷. Asymmetric cell divisions are also common at later stages of development in mammals⁵⁶. To investigate the effect of this mechanism of commitment on quantitative fate mapping, we simulated barcoding experiments on a pectinate 16 terminal type fate map (Figure 1C, bottom left) with commitments happening with either asymmetric divisions or symmetric division exclusively. The pectinate fate map with 16 terminal types was modified slightly for the comparison of commitment via symmetric and asymmetric cell division. All progenitor state biases were set to 0.5. To obtain a fate map whose terminal types have similar population sizes, the doubling time for the terminal state that was the i -th to emerge was set to $0.6 - 0.5/16i$. Moreover, the cell death probabilities for the terminal population were set to 0.1 and a cell doubling probability per doubling time ($c_{T_j,L}$ for terminal type T_j) of 0.4 was instituted. For the asymmetric mode, one daughter cell commits to the downstream terminal type and the other commits to the downstream progenitor state after commitment (Figure S5T). For the symmetric mode, each cell randomly commits to one of the downstream fates based on predefined probabilities and subsequently undergoes symmetric divisions (similar to all previous simulations) (Figures 1 and S5T). In each case, we simulated one hundred experiments with 50 hgRNAs and 50 cells sampled from each terminal type (Figures S5U,V). We then applied Phylotime and ICE-FASE to obtain quantitative fate maps and compared the results between the two commitment modes. For the results and detailed interpretation, see Figures S5T–Z and their legends.

ICE-FASE for multiple replicates—First, the FASE distance matrix was computed for each phylogeny separately. The distance matrices were then averaged, and a consensus topology was obtained by applying UPGMA clustering to the average FASE distance matrix. The set of inferred progenitor states were defined based on the consensus topology. For each replicate's phylogeny, each of its nodes were assigned one of the progenitor states based on their potencies, based on which ICEs were identified. Next, ICEs identified from all replicates were pooled, and their average time was taken to be the consensus commitment time. Finally, edges associated with each inferred progenitor state were identified in each replicate's phylogeny based on its consensus commitment time. The resulting population sizes of the progenitor states, and population sizes committing to each downstream states or types were averaged. The average population size was taken as the consensus population size and the proportion of average population size committing to each downstream state was taken as the consensus commitment bias.

Simulation of experiments with prolonged commitment—Commitment events modeled thus far involve all the progenitor population committing within one cell cycle. Because some progenitor states undergo gradual commitment during development, we expanded our models to allow progenitor states to commit over more extended time periods. We then created three fate maps with 16 terminal types that are identical in topology and

other parameters, except that P11 commits to its downstream states (P1 and P9) between day 5.8 and 6.3 in the first map, between day 5.8 and 6.9 in the second map, and between 5.8 and 7.5 days in the third map (Figure S6A). The details of the modifications are given in the next section. We simulated phylogenies based on these fate maps with fixed sampling of 100 cells from each terminal type and repeated simulations one hundred times for each. After applying Phylotime followed by the ICE-FASE algorithm, we compared the accuracy of quantitative fate map reconstruction in each case. For the results and detailed interpretations, see Figures S6B–E and their legends.

Modified fate map with prolonged commitment—To illustrate the effect of progenitor cells committing over longer times, we modified an existing fate map with 16 terminal types from our panel. In the original fate map, the progenitor state P11 commits within a single cell cycle from 5.8 and 6.3 days. In a first modified version, the commitment happens within two cell cycles from 5.8 and 6.9 days. A random $\frac{1}{3}$ of all the P11 population commits at the original time whereas the remaining population self-renews by going through an extra round of cell division. After the extra round of self-renewal, they commit to the downstream cell states of P9 and P1 according to the P11 commitment bias. Finally, in a second modified version, commitment happens over three cell cycles from 5.8 and 7.5 days. A random third of all the progenitor population again commit at the original commitment time. After one extra round of self-renewal, half of the remaining uncommitted population commit while the other half self-renew for yet another round, after which they also commit. To accommodate commitments happening over more than one cell cycle, the definition of commitment time and progenitor population size were modified accordingly. The progenitor population size was defined as the sum of the population size committing to downstream states during each cell cycle. The commitment time was defined as the start of the cell cycle during which commitment has happened weighted by the number of cells that have committed during the cell cycle. The ICE-FASE estimates of P11 for each fate map were compared to the respective definition of commitment time and progenitor population size.

Simulation of experiments with multifurcating fate commitments—Commitment events modeled thus far involved bifurcating progenitor states. However, some progenitor states in development may simultaneously commit to more than two immediate downstream states (e.g., trifurcate). To investigate this, our generative model was extended to simulate n-furcations in general. The details of the model extension are given in the next section. Next, we assessed whether the ICE-FASE algorithm can resolve trifurcations. First, a trifurcation in a 16 terminal type fate map was introduced by removing the progenitor state P8 from the fate map and making its downstream states (P5 and P6) directly emerge from the commitment event of its upstream state P11 (Figure S6F). We refer to this trifurcating progenitor state as P11*, compared to the original bifurcating P11 as they have identical potencies, commitment times, and population sizes. We then simulated phylogenies and barcoding with 50 hgRNAs based on both the original bifurcating and the new trifurcating fate maps with fixed sampling of 100 cells from each terminal type and repeated each simulation one hundred times. Finally, we applied Phylotime followed by the ICE-FASE algorithm to reconstruct a quantitative fate map based on each simulated experiment. For the results and detailed interpretation see Figures S6G–L and their legends.

Count graph for simulating progenitor state multifurcation—When the fate map involved a multifurcation, the model to generate cell phylogeny was extended. Only multifurcation via symmetric cell division was considered, where only minor adjustment to the methods described earlier needed to be made. Take a trifurcation of progenitor state P_i into P_j , P_k and P_l as an example. All computations with regard to process 1 (proliferate process) remained the same. During process 2 (commit and proliferate process) in the count graph generation step, the default count node was split into three sub-count nodes of commitment modes $P_i-P_jP_j$, $P_i-P_kP_k$ and $P_i-P_lP_l$. The sample size generation step was carried out as usual. In the reorganization step, six different new sub-count nodes were created to replace the original three, including $P_i-P_jP_j$, $P_i-P_kP_k$, $P_i-P_lP_l$, P_i-P_j , P_i-P_k and P_i-P_l . During reorganization, non-coalesced cells in $P_i-P_jP_j$, $P_i-P_kP_k$, $P_i-P_lP_l$ are moved to their respective non-coalesced sub-count nodes. The remaining steps were carried out as usual.

Generation of an inducible Cas9 barcoded stem cell line

Knock-in of an Inducible Cas9 Cassette: EPI iPSCs were modified to express Cas9 protein under doxycycline induction. CRISPR/Cas9 was used to target and insert both a reverse tetracycline-controlled transactivator (rtTA) construct and a tetracycline-dependent Cas9 construct into each of the two copies of the AAVS1 safe harbor locus. To stably introduce the cassette, cells were grown to 80% confluency and then dissociated with Accutase for 13 minutes to generate a single cell suspension. Dissociated cells were resuspended in mTeSR Plus media with 5 μ M blebbistatin and counted- 50,000 cells were seeded into one well of a 24-well plate coated in Matrigel. The following day, 350 ng of plasmid expressing Cas9 and an AAVS1-targeted guide RNA (modified PX459, with T2A replaced by P2A) and 500 ng each of plasmids containing the Cas9 donor sequence (modified Addgene #58409 with blasticidin resistance)⁴⁷ and rtTA donor sequence (Addgene #60843)⁵⁰ were combined and added to 48 μ l Opti-MEM (ThermoFisher). 2 μ l of Lipofectamine Stem Transfection Reagent (ThermoFisher) were added to the transfection mix, which was then vortexed and incubated for ten minutes at room temperature. The entire transfection mix was added to one well of cells. Media was replaced the following day. 40 hours after transfection, cells were transiently selected for 24 hours with 0.95 μ g/mL puromycin (MilliporeSigma), 5 μ g/mL blasticidin (MilliporeSigma), and 200 μ g/mL G418 sulfate (ThermoFisher).

Surviving cells were cultured to 30% confluency, and then dissociated to a single cell suspension for clonal expansion. 500–1000 cells were seeded in one well of a 6-well plate and cultured for 7–10 days before clonal colonies were picked and screened for the intended insertions. PCR was performed with a reverse primer complementary to the right homology arm of the targeted AAVS1 locus (GGAACGGGGCTCAGTCTGA) and a forward primer either targeting the Cas9 knockin (CACCTTGACTCGTCGGTGA) or the rtTA constructs knockin (GCTGATTATGATCCTGCAAGC). Positive colonies were cultured and clonally expanded once more, with a second round of colony picking and PCR screening to ensure clonality of the final cell line.

Selected clones were then further screened for functionality of the inducible Cas9 cassette. Cells were treated with 2 ug/mL doxycycline (MilliporeSigma) for 5 days and cells were harvested for RNA extraction. RNA was converted into cDNA and qPCR was performed to confirm both *Cas9* and *rtTA* mRNA expression when normalized against housekeeping genes *GAPDH*, *CREBBP*, and *ACTB* (*Cas9* Fwd: CCGAAGAGGTCGTGAAGAAG; *Cas9* Rev: GCCTTATCCAGTTCGCTCAG; *rtTA* Fwd: GCTAAAGTGCATCTCGGCAC; *rtTA* Rev: TGTTCCTCCAATACGCAGCC; *GAPDH* Fwd: TAGCCAAATTCGTTGTCATACC; *GAPDH* Rev: CTGACTTCAACAGCGACACC; *CREBBP* Fwd: GAGAGCAAGCAAACGGAGAG; *CREBBP* Rev: AAGGGAGGCAAACAGGACA; *ACTB* Fwd: GCGAGAAGATGACCCAGATC; *ACTB* Rev: CCAGTGGTACGGCCAGAGG).

Transfection with a homing guide RNA library: 50,000 cells from the doxycycline-inducible Cas9 line were seeded into one well of a 24-well plate. The following day, 300 ng of Super PiggyBac Transposase (SBI System Biosciences), 700 ng of PB-U6insert hgRNA library (Addgene #104536)⁵, and 50 ng of PB-U6insert-EF1puro library (Addgene #104537)⁵ were combined and added to 48 ul Opti-MEM (Gibco). 2 ul of Lipofectamine Stem Transfection Reagent (Thermo Fisher) were added to the transfection mix, which was then vortexed and incubated for ten minutes at room temperature. The entire transfection mix was added to one well of cells. Media was replaced the following day. Transfected cells were selected with 0.95 ug/mL puromycin for one week.

Selected cells were dissociated to single cell suspensions, and 500–1000 cells were seeded in one well of a 6-well plate and cultured for 7–10 days before clonal colonies were picked. Colonies were screened for relatively high numbers of hgRNA insertions using qPCR. Genomic DNA was extracted from each colony and the relative number of hgRNA insertions was measured by subtracting Cq values of genomic hgRNA amplification (Fwd: ATGGACTATCATATGCTTACCGT; Rev: TTCAAGTTGATAACGGACTAGC) from Cq values of genomic *SOX11* amplification (Fwd: TGATGTTGACCTGAGCTTG; Rev: TAGTCGGGAACTCGAAGTG). Colonies with the largest cycle threshold value difference, indicating the highest number of hgRNA insertions, were cultured and clonally expanded once more. Colonies were picked one additional time to ensure clonality of the final EP1-Cas9-hgRNA iPS cell line.

Barcoding activity of cell line-integrated hgRNAs: Cell line hgRNA identifiers and mutation activity levels were determined by performing a doxycycline time-course experiment- cells were treated with 1 ug/mL doxycycline daily for 11 days. Genomic DNA was extracted from cells after 0, 4, 8 and 11 days of doxycycline treatment and Cas9 induction. hgRNA sequencing libraries were prepared as follows: 5 ng of genomic DNA was amplified on a real-time PCR machine in 1X KAPA SYBR FAST (Roche) with 0.0625 uM hgRNA PCR1 forward primer (ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGGACTATCATATGCTTACCGT), 0.1875 uM hgRNA PCR1 truncated forward primer (CTACACTCTTTCCCTACACGAC), 0.0625 uM hgRNA PCR1 reverse primer (TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCATACCAATGGCCCGAA),

and 0.1875 μ M hgRNA PCR1 truncated reverse primer (GTGACTGGAGTTCAGACGTG). Reactions were denatured at 95°C for three minutes, and then cycled at 95°C for 20 seconds, 64°C for 20 seconds, and 72°C for ten seconds. The reactions were stopped when the real-time PCR curve reached early-to-mid-exponential phase. Reactions were then diluted 10–100 fold. 1 μ l of diluted PCR was used as template for a subsequent PCR reaction in 1X KAPA HiFi HotStart ReadyMix (Roche) with 1X SYBR Green I Nucleic Acid Stain (ThermoFisher) and 0.25 μ M each dual indexing primer pair for Illumina. Reactions were denatured at 98°C for thirty seconds, and then cycled at 98°C for 10 seconds, 64°C for 20 seconds, and 72°C for ten seconds. Reactions were stopped when the real-time PCR curve reached early-to-mid-exponential phase. Libraries were purified with DNA Clean & Concentrator-5 columns (Zymo), sequenced on an Illumina MiSeq instrument, and analyzed. hgRNA sequencing reads were analyzed using the published MARC1 data analysis pipeline³⁰. The percent of reads for each hgRNA identifier sequence that were mutated was calculated at each time point (Figure S4), determining the relative activity for every hgRNA.

In vitro quantitative fate map experiments—Single cells from the EP1-Cas9-hgRNA iPSC cell line were FACS sorted into a 96-well plate coated with Matrigel and containing mTeSR plus medium supplemented with 5 μ M blebbistatin, 10% CloneR (STEMCELL Technologies), 1 μ M Pifithrin- α hydrobromide (Tocris Bioscience), 1X Antibiotic-Antimycotic (ThermoFisher), and 0.2 μ g/uL doxycycline. Supplemented media was exchanged every other day. Three days after sorting, wells with surviving cells were identified and assigned to follow the quantitative fate map for either experiment one (E1) or experiment two (E2).

For E1, cells were grown in the sorted wells until there were approximately 32 cells, or 6 doublings, which was considered to be P5 of the QFM. P5 cells were then carefully dissociated and passaged—media was aspirated from wells and 30 μ l of Accutase was added and incubated at 37°C for 8 minutes. The Accutase was then gently triturated to detach all cells from the plate, and 10 μ l was directly added to a new well filled with supplemented media, corresponding to P3 of the QFM. The remaining 20 μ l was directly added to a new well corresponding to P4, resulting in a 1:2 split of the cells. Passaged cells were incubated for 2 hours to allow cells to settle and attach to the Matrigel coating, after which a 50% media exchange was performed to decrease the total amount of Accutase remaining in the wells. Once the cells in P4 had gone through approximately two doublings, they were passed following the same protocol, this time splitting the cells evenly between progenitor states P1 and P2. After P3's cells had gone through approximately 4 doublings since their passage they were split once again, evenly into terminal types T5 and T6. P2 cells were given four doublings before being split evenly into terminal types T3 and T4, and finally P1 was split evenly after five doublings into terminal types T1 and T2.

For E2, cells were propagated using the same techniques as E1, but following a slightly altered QFM. P3 cells were split into terminal types T5 and T6 after just two doublings, and P4 cells were split into P1 and P2 after four doublings, representing a switch in the order of commitment P3 and P4 commitment times between the E1 and E2 fate maps. After cells were passaged into their terminal wells, doxycycline treatment was ended so that

barcode editing would discontinue. Cells were then passaged into 6-well plates and grown to confluency for terminal cell barcode extraction and analysis.

Sequencing single-cell lineage barcodes—Each terminal well from E1 and E2 were dissociated into single cell suspensions by incubating the cells in accutase for 14 minutes. Single cells were resuspended and diluted in PBS pH 7.4 for FACs sorting. Single cells from each terminal group were sorted into 192 wells of a 384-well plate (half a plate per terminal group), with each well containing 1 ul of QuickExtract DNA Extraction Buffer (Lucigen). Plates were vortexed and spun down directly after sorting to ensure cells were in the QuickExtract buffer. DNA was extracted from single cells by incubating the sorted plates for 10 minutes at 65°C followed by 5 minutes at 98°C to inactivate the QuickExtract.

Single-cell hgRNA sequencing libraries were generated using three serial PCR reactions as per the published protocol from Leeper et al.³⁰, with each cell treated as an individual sample. Each well of single-cell DNA received 9 ul of PCR0 mix containing 1X DreamTaq Hot Start PCR Master Mix (ThermoFisher) and 0.5 uM each hgRNA pre-amplification forward and reverse primers (Fwd: AAGTAATAATTTCTTGGGTAGTTTGCAG; Rev: GAAAAAGCCATACCAATGGGC). Reactions were denatured at 95°C for three minutes, cycled five times for 95°C for 20 seconds, 55°C for 30 seconds, and 72°C for one minute, and then cycled 20 times for 95°C for 20 seconds, 60°C for 30 seconds, and 72°C for 30 seconds.

One microliter of each cell's PCR0 reaction was used as template for the proceeding PCR1 reaction. Reactions for PCR1 and 2 were carried out as described above in the section "Determining profiles for integrated hgRNAs", with each single cell continuing to be amplified and indexed as an individual sample. After PCR2, the reactions for each terminal cell type were normalized and purified together to create one combined sequencing library per terminal group. Libraries were quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher) sequenced on a MiSeq System using Miseq Reagent Micro Kits (Illumina).

Determining actual progenitor population size from in vitro experiments—For each experiment, brightfield images were taken of P5, P4, and P3 wells just before passaging (Figure S6). To estimate the cell numbers at each progenitor state, images were analyzed in ImageJ. Outlines were drawn around 5 different cells within a colony and their average area was measured. All colonies within a well were then outlined and the total combined area of the colonies was measured and divided by the average cell area of a cell to estimate the total number of cells present in each well.

Processing of in vitro experimental data—Identifier and spacer sequence pairs were extracted for each sample using the initial step (BLAST search) detailed in the published pipeline³⁰. For cells that were sequenced more than once, pair counts for unique "identifier+spacer" were first merged for each cell. The merged data were then provided as input to the remainder of the pipelines for sequencing error correction and filtering.

A total of 32 hgRNAs were identified from the filtered results, each observed in more than 921 cells. Sequencing errors among identifier sequences were first corrected. For each

identifier sequence that was not one of the 32 hgRNAs observed in the unmutated sample, if the identifier sequence was within a hamming distance of 1 to any true identifiers, its spacer counts were merged with that of the known hgRNA's. After the correction, no other identifiers other than the 32 known hgRNAs were observed in more than 3 cells.

Spacer sequencing errors were corrected next. First, the error reads within each cell and hgRNA combination were corrected. In one of our sequencing runs, one cycle of sequencing returned 'N' for all spacer sequences. These errors were computationally corrected: if there existed another spacer sequence for the same identifier and cell that was exactly the same except for the 'N' base pair, the count of the error spacer with the 'N' base was merged with the other spacer. Next, the spacer sequencing errors across different cells were corrected. Again, the error involving 'N' base pairs were further corrected across the cells using the same criteria as the within cell correction.

Each allele was labeled as unmutated or mutated by comparing the spacer sequences to that of the reference sequencing result, that is, if a spacer sequence was observed in the parent for the same identifier, it was labeled as unmutated. One identifier "GCCAAAAGCT" did not amplify in the parent data, and the sequence "GAAACACCGGTGGTCGCCGTGGAGAGTGGTGGGGTTAGAGCTAGAAATAG" was identified as the unmutated spacer based on alignments of its different observed alleles.

Noisy reads were further filtered for cell+hgRNA combinations that had more than one spacer observed. If the most abundant spacer was at least four times more abundant than all the other spacer reads observed, only the most abundant spacer was kept. All the spacer counts with fewer than five total reads were also excluded.

After processing, 1197 cell+hgRNA combinations out of the total 54,012 (2.2%) still had more than one spacer observed. If more than two spacers were observed for more than two hgRNAs in a single cell, the cell was likely a doublet. 33 such cells were identified and filtered. Each cell had a median 25 out of 32 hgRNAs detected. Each hgRNA was detected in a median of 83% of all cells (Figure S7D). Before reconstruction, non-informative cells and hgRNAs were filtered out. Any hgRNA with a diversity of one, that is, all cells in which an hgRNA was observed had the same allele, was considered non-informative and was excluded. An allele was considered informative if it was mutated and observed in more than one cell in each group, and cells with less than three informative alleles were filtered out. In all, 970 out of 1051 cells for 31 hgRNAs passed the filters for E1 and 943 out of 1032 cells for 29 hgRNAs passed filters for E2. For Phylotime reconstruction, naïve estimates for mutation rate and uniform prior for mutant allele emergence probabilities were used.

Simulation and ground truth fate map of in vitro experiment—To conduct simulations that best resembled the in vitro experiment, the effective cell division rates during the experiment were first determined. First, the division rate of P5 was chosen so that the population size at the first split is the most consistent with what was observed. Next, we assumed that cells were split in proportion to the volume of the suspension as P5 was split into P3 and P4. The division rate of P3 and P4 were set so that their respective population sizes at the split agreed with what was observed. The division rate for P1, P2 and

all the terminal cells were set to once every 20 hours. The exact division rate chosen and the population size at each stage are detailed in Figure S7B. Notice that the commitment happens one cell division prior to the well split, so the ground truth commitment time is one cell division earlier than the spit time, and the progenitor field size is half of what was observed in the well at the split.

To simulate hgRNA barcodes from ground truth fate map, mutation rates were estimated from the time course data of bulk mutated fractions from the iPSC line. For mutant allele emergence probabilities, predictions from inDelphi were used.

QUANTIFICATION AND STATISTICAL ANALYSIS

Estimation of mutagenesis parameters in MARC1 mice—To get posterior estimates of mutation rates of MARC1 hgRNAs (i.e., λ , rate of the Poisson process), a grid search was conducted to match empirical distributions of mutated fractions among simulated and observed data across several embryonic time points. Previously reported hgRNAs formed three classes: the ‘slow’ class generated mutations on the order of 0.001 mutations/day, ‘intermediate’ class generated ~ 0.1 mutations/days, and fast class generated ~ 1.0 mutations/day during early mouse development. The ‘slow’ and ‘fast’ estimates expectedly had large uncertainties as most observed fractions are close to 0 or 100 percent mutated. Alternatively, a naïve estimate of mutation rate can also be used. If mutated fractions F_i were observed at time T_i in animal i for $i = 1, \dots, N$, then $\hat{\lambda} = \frac{1}{N} \sum_{i=1}^N \frac{1}{T_i} \log(1 - F_i)$ is a naïve estimate.

For mutant alleles of a barcoding site, estimating the probabilities of individual repair outcomes created by Cas9 DNA break-repair (mutant emergence probabilities) was challenging. Normally, the fraction of cells carrying a particular mutant allele among all cells with a mutated allele (within-animal estimates) is a good estimator of the allele emergence probabilities. However, when cells divide and mutate starting from a small field size, these fractions are largely affected by the time of the mutagenesis events, as early events result in larger clones carrying the same exact mutation. On the other hand, when hgRNA genotypes are observed for multiple animals, the fraction of animals that carry a particular genotype, once normalized, and when the probability is small, can be good estimates to the mutation probabilities (across-animal estimates) (Figures S4B,C). In this case, the estimation accuracy depends on the number of animals analyzed. From the MARC1 time course data, the within-animal estimates were calculated for each animal and averaged, and the across-animal estimates were calculated based on 173 embryos from 2 mouse lines. To get a more complete profile of possible mutant alleles and their occurrence probabilities for each hgRNA, we adapted the inDelphi machine learning algorithm to predict CRISPR-Cas9 mutation results³¹ for hgRNAs. We observed that the inDelphi-predicted probabilities agreed well with the across-animal estimates from MARC1, but poorly with within-animal estimates (Figure S4B). Further, the fact that the majority of the low probability mutations were not observed in any mouse suggests that the limited number of hgRNA mutation events during mouse development does not sufficiently cover a large portion of the mutational profiles. These conclusions were further validated by simulating multiple animal lineage barcode data based on inDelphi-predicted mutational

profiles and comparing the within- and across-animal estimates from the simulated data of the true parameters (Figure S4C).

InDelphi predictions of hgRNA allele emergence probabilities—The emergence probabilities of hgRNA mutant alleles were computed by inDelphi. inDelphi is a machine learning algorithm to predict heterogeneous insertions and deletions resulting from CRISPR/Cas9 double-strand break³¹. In this study, inDelphi model trained with the mouse embryonic stem cell mutation dataset was used to predict the probabilities of hgRNA mutants from MARC1 mice. The original 64 hgRNA sequences in MARC1 mice were used as inputs. Since Cas9 nuclease cuts 3 bp upstream of the Protospacer Adjacent Motif (PAM, NGG sequence)¹, the possible mutations from the cut site at –3 bp from the PAM sequence were computed. To take into account the repeated targeting of hgRNAs, inDelphi is first applied to predict a set of first-round mutations. Subsequently, the resulting first-round mutations were used as inputs to the next round of inDelphi predictions. Notably, only mutant sequences with >16 bp protospacer and PAM were subject to the second-round analysis as gRNA without >16 bp spacer sequence loses its activity⁵⁷. Here, the probabilities of the next generation mutants were computed by multiplying the probabilities of the mutant in the current round by the probabilities of the mutant in the previous round. Repetitive application of inDelphi produces exponentially growing numbers of potential mutant alleles. Therefore, the analysis was limited to three cycles, resulting in first to third generations of mutants. The same mutation can be created in multiple rounds, in such cases, the probabilities from multiple rounds were summed. Finally, probabilities of all mutant alleles were normalized to have a sum of one. The inDelphi predictions for hgRNAs are provided on Zenodo (see Key Resources Table).

Assessing accuracy of progenitor state parameter estimates with RMSE—As a measurement of accuracy for the progenitor state parameter estimation, that is commitment time, progenitor population size and commitment bias, the root mean square error (RMSE) was used to quantify the amount of error in parameter estimation. The root mean square error is defined as

$$RMSE = \sqrt{\frac{1}{I} \sum_i^I (\hat{x}_i - x_i)^2}$$

where the i index is over the number of progenitor states in the reconstructed fate map, x_i is the true parameter and \hat{x}_i is its estimate. For progenitor population size, RMSE was computed for the log₂ transformed population sizes. In cases where the reconstructed topology is used to assess parameter estimation accuracies, only the inferred progenitor states that could be correctly mapped to the truth were included. In cases where the true topology was used to assess parameter estimation accuracies, the RMSE is over all the progenitor states. Note that combining all progenitor states irrespective of sampling fraction inflates the RMSE values.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge Kian Kalhor, Dr. Yuxin Zhu, Dr. Justus Kebschull, and Dr. Loyal Goff for comments on the manuscript. This work was supported by grants from the Simons Foundation (SFARI 606178, R.K.) and the National Institutes of Health (NIH) (U01HL156056, R.K.; R01HG012357, R.K.; R01HG009518, H.J.; R01HG010889, H.J.; P30EY001765, D.J.Z.; F31EY030769, C.M.B.) and the David & Lucile Packard Foundation (2020-71380, R.K.). Computation was carried out at the Advanced Research Computing at Hopkins (ARCH), supported by the National Science Foundation (NSF) (OAC1920103). D.J.Z. is supported by Research to Prevent Blindness and the Guerrieri Family Foundation.

References

1. Stadler T, Pybus OG, and Stumpf MPH (2021). Phylodynamics for cell biologists. *Science* 371, eaah6266. 10.1126/science.aah6266. [PubMed: 33446527]
2. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, and van Oudenaarden A (2018). Whole-organism clone tracing using single-cell sequencing. *Nature* 556, 108–112. 10.1038/nature25969. [PubMed: 29590089]
3. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, and Shendure J (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353, aaf7907. 10.1126/science.aaf7907. [PubMed: 27229144]
4. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, and Junker JP (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol* 36, 469–473. 10.1038/nbt.4124. [PubMed: 29644996]
5. Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, and Church GM (2018). Developmental barcoding of whole mouse via homing CRISPR. *Science* 361, 927–936. 10.1126/science.aat9804.
6. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, Jost M, Quinn JJ, Yang D, Jones MG, et al. (2019). Molecular recording of mammalian embryogenesis. *Nature* 570, 77–82. 10.1038/s41586-019-1184-5. [PubMed: 31086336]
7. Bowling S, Sritharan D, Osorio FG, Nguyen M, Cheung P, Rodriguez-Fraticelli A, Patel S, Yuan W-C, Fujiwara Y, Li BE, et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* 181, 1410–1422.e27. 10.1016/j.cell.2020.04.048. [PubMed: 32413320]
8. Bizzotto S, Dou Y, Ganz J, Doan RN, Kwon M, Bohrsen CL, Kim SN, Bae T, Abyzov A, NIMH Brain Somatic Mosaicism Network, et al. (2021). Landmarks of human embryonic development inscribed in somatic mutations. *Science* 371, 1249–1253. 10.1126/science.abe1544. [PubMed: 33737485]
9. Spencer Chapman M, Ranzoni AM, Myers B, Williams N, Coorens THH, Mitchell E, Butler T, Dawson KJ, Hooks Y, Moore L, et al. (2021). Lineage tracing of human development through somatic mutations. *Nature* 595, 85–90. 10.1038/s41586-021-03548-6. [PubMed: 33981037]
10. Baron CS, and van Oudenaarden A (2019). Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol* 20, 753–765. 10.1038/s41580-019-0186-3. [PubMed: 31690888]
11. Sulston JE, Schierenberg E, White JG, and Thomson JN (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol* 100, 64–119. [PubMed: 6684600]
12. Salipante SJ, and Horwitz MS (2006). Phylogenetic fate mapping. *Proc. Natl. Acad. Sci. U. S. A* 103, 5448–5453. 10.1073/pnas.0601265103. [PubMed: 16569691]
13. Wagner DE, and Klein AM (2020). Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet* 10.1038/s41576-020-0223-2.
14. Coorens THH, Moore L, Robinson PS, Sanghvi R, Christopher J, Hewinson J, Przybilla MJ, Lawson ARJ, Spencer Chapman M, Cagan A, et al. (2021). Extensive phylogenies of human development inferred from somatic mutations. *Nature* 597, 387–392. 10.1038/s41586-021-03790-y. [PubMed: 34433963]
15. Weinreb C, and Klein AM (2020). Lineage reconstruction from clonal correlations. *Proc. Natl. Acad. Sci. U. S. A* 117, 17041–17048. 10.1073/pnas.2000238117. [PubMed: 32632001]

16. Salvador-Martínez I, Grillo M, Averof M, and Telford MJ (2019). Is it possible to reconstruct an accurate cell lineage using CRISPR recorders? *Elife* 8. 10.7554/eLife.40292.
17. Chor B, and Tuller T (2005). Maximum likelihood of evolutionary trees: hardness and approximation. *Bioinformatics* 21 Suppl 1, i97–106. 10.1093/bioinformatics/bti1027. [PubMed: 15961504]
18. Ebisuya M, and Briscoe J (2018). What does time mean in development? *Development* 145. 10.1242/dev.164368.
19. Lui JH, Hansen DV, and Kriegstein AR (2011). Development and evolution of the human neocortex. *Cell* 146, 18–36. 10.1016/j.cell.2011.06.030. [PubMed: 21729779]
20. Leung CY, and Zernicka-Goetz M (2015). Mapping the journey from totipotency to lineage specification in the mouse embryo. *Curr. Opin. Genet. Dev* 34, 71–76. 10.1016/j.gde.2015.08.002. [PubMed: 26343010]
21. Davidson EH (1993). Later embryogenesis: regulatory circuitry in morphogenetic fields. *Development* 118, 665–690. 10.1242/dev.118.3.665. [PubMed: 7915668]
22. Shao K-T, and Sokal RR (1990). Tree Balance. *Syst. Biol* 39, 266–276. 10.2307/2992186.
23. Ciemerych MA, and Sicinski P (2005). Cell cycle in mouse development. *Oncogene* 24, 2877–2898. 10.1038/sj.onc.1208608. [PubMed: 15838522]
24. Fabian D, Makarevich AV, Chrenek P, Bukovská A, and Koppel J (2007). Chronological appearance of spontaneous and induced apoptosis during preimplantation development of rabbit and mouse embryos. *Theriogenology* 68, 1271–1281. 10.1016/j.theriogenology.2007.08.025. [PubMed: 17915306]
25. Kingman JFC (1982). On the genealogy of large populations. *J. Appl. Probab* 19, 27–43. DOI: 10.2307/3213548.
26. Kingman JFC (1982). The coalescent. *Stoch. Process. their Appl* 13, 235–248. 10.1016/0304-4149(82)90011-4.
27. Yang Z (2014). *Molecular Evolution: A Statistical Approach* (Oxford University Press).
28. Kendall M, and Colijn C (2016). Mapping Phylogenetic Trees to Reveal Distinct Patterns of Evolution. *Mol. Biol. Evol* 33, 2735–2743. 10.1093/molbev/msw124. [PubMed: 27343287]
29. Kalhor R, Mali P, and Church GM (2017). Rapidly evolving homing CRISPR barcodes. *Nat. Methods* 14, 195–200. 10.1038/nmeth.4108. [PubMed: 27918539]
30. Leeper K, Kalhor K, Vernet A, Graveline A, Church GM, Mali P, and Kalhor R (2021). Lineage barcoding in mice with homing CRISPR. *Nat. Protoc* 16, 2088–2108. 10.1038/s41596-020-00485-y. [PubMed: 33692551]
31. Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, Cassa CA, Liu DR, Gifford DK, and Sherwood RI (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 563, 646–651. 10.1038/s41586-018-0686-x. [PubMed: 30405244]
32. Feng J, Dewitt WS 3rd, McKenna A, Simon N, Willis AD, and Matsen FA 4th (2021). ESTIMATION OF CELL LINEAGE TREES BY MAXIMUM-LIKELIHOOD PHYLOGENETICS. *Ann. Appl. Stat* 15, 343–362. 10.1214/20-aos1400. [PubMed: 35990087]
33. Gong W, Granados AA, Hu J, Jones MG, Raz O, Salvador-Martínez I, Zhang H, Chow K-HK, Kwak I-Y, Retkute R, et al. (2021). Benchmarked approaches for reconstruction of in vitro cell lineages and in silico models of *C. elegans* and *M. musculus* developmental trees. *Cell Syst.* 12, 810–826.e4. 10.1016/j.cels.2021.05.008. [PubMed: 34146472]
34. Seidel S, and Stadler T (2022). TiDeTree: A Bayesian phylogenetic framework to estimate single-cell trees and population dynamic parameters from genetic lineage tracing data. *bioRxiv*, 2022.02.14.480422 10.1101/2022.02.14.480422.
35. Jones MG, Khodaverdian A, Quinn JJ, Chan MM, Hussmann JA, Wang R, Xu C, Weissman JS, and Yosef N (2020). Inference of single-cell phylogenies from lineage tracing data using *Cassiopeia*. *Genome Biol.* 21, 92. 10.1186/s13059-020-02000-8. [PubMed: 32290857]
36. Conradt B (2009). Genetic control of programmed cell death during animal development. *Annu. Rev. Genet* 43, 493–523. 10.1146/annurev.genet.42.110807.091533. [PubMed: 19886811]
37. Konstantinides N, Holguera I, Rossi AM, Escobar A, Dudragne L, Chen Y-C, Tran TN, Martínez Jaimes AM, Özel MN, Simon F, et al. (2022). A complete temporal transcription factor series in the fly visual system. *Nature* 604, 316–322. 10.1038/s41586-022-04564-w. [PubMed: 35388222]

38. Hormoz S, Singer ZS, Linton JM, Antebi YE, Shraiman BI, and Elowitz MB (2016). Inferring Cell-State Transition Dynamics from Lineage Trees and Endpoint Single-Cell Measurements. *Cell Syst.* 3, 419–433.e8. 10.1016/j.cels.2016.10.015. [PubMed: 27883889]
39. Forrow A, and Schiebinger G (2021). LineageOT is a unified framework for lineage tracing and trajectory inference. *Nat. Commun* 12, 4940. 10.1038/s41467-021-25133-1. [PubMed: 34400634]
40. Yao Z, Liu K, Deng S, and He X (2021). An instantaneous coalescent method insensitive to population structure. *J. Genet. Genomics* 48, 219–224. 10.1016/j.jgg.2021.03.005. [PubMed: 34001433]
41. Mittnenzweig M, Maysar Y, Cheng S, Ben-Yair R, Hadas R, Rais Y, Chomsky E, Reines N, Uzonyi A, Lumerman L, et al. (2021). A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* 184, 2825–2842.e22. 10.1016/j.cell.2021.04.004. [PubMed: 33932341]
42. Cagan A, Baez-Ortega A, Brzozowska N, Abascal F, Coorens THH, Sanders MA, Lawson ARJ, Harvey LMR, Bhosle S, Jones D, et al. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature* 604. 10.1038/s41586-022-04618-z.
43. Kojima Y, Tam OH, and Tam PPL (2014). Timing of developmental events in the early mouse embryo. *Semin. Cell Dev. Biol* 34, 65–75. 10.1016/j.semcdb.2014.06.010. [PubMed: 24954643]
44. Ford D, Matsen FA, and Stadler T (2009). A method for investigating relative timing information on phylogenetic trees. *Syst. Biol* 58, 167–183. 10.1093/sysbio/syp018. [PubMed: 20525576]
45. Eldred KC, Hadyniak SE, Hussey KA, Brenerman B, Zhang P-W, Chamling X, Sluch VM, Welsbie DS, Hattar S, Taylor J, et al. (2018). Thyroid hormone signaling specifies cone subtypes in human retinal organoids. *Science* 362. 10.1126/science.aau6348.
46. Wahlin KJ, Maruotti JA, Sripathi SR, Ball J, Angueyra JM, Kim C, Grebe R, Li W, Jones BW, and Zack DJ (2017). Photoreceptor Outer Segment-like Structures in Long-Term 3D Retinas from Human Pluripotent Stem Cells. *Sci. Rep* 7, 766. 10.1038/s41598-017-00774-9. [PubMed: 28396597]
47. González F, Zhu Z, Shi Z-D, Lelli K, Verma N, Li QV, and Huangfu D (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* 15, 215–226. 10.1016/j.stem.2014.05.018. [PubMed: 24931489]
48. Maruotti J, Wahlin K, Gorrell D, Bhutto I, Luty G, and Zack DJ (2013). A simple and scalable process for the differentiation of retinal pigment epithelium from human pluripotent stem cells. *Stem Cells Transl. Med* 2, 341–354. 10.5966/sctm.2012-0106. [PubMed: 23585288]
49. Remeseiro S, Cuadrado A, Gómez-López G, Pisano DG, and Losada A (2012). A unique role of cohesin-SA1 in gene regulation and development. *EMBO J.* 31, 2090–2102. 10.1038/emboj.2012.60. [PubMed: 22415368]
50. DeKolver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijnsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, et al. (2010). Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* 20, 1133–1142. 10.1101/gr.106773.110. [PubMed: 20508142]
51. Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. 10.1038/nmeth.2089. [PubMed: 22930834]
52. Bhise NS, Wahlin KJ, Zack DJ, and Green JJ (2013). Evaluating the potential of poly(beta-amino ester) nanoparticles for reprogramming human fibroblasts to become induced pluripotent stem cells. *Int. J. Nanomedicine* 8, 4641–4658. 10.2147/IJN.S53830. [PubMed: 24348039]
53. Paradis E, and Schliep K (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. 10.1093/bioinformatics/bty633. [PubMed: 30016406]
54. Schliep KP (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. 10.1093/bioinformatics/btq706. [PubMed: 21169378]
55. Chen T, and Guestrin C (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '16*. (Association for Computing Machinery), pp. 785–794. 10.1145/2939672.2939785.
56. Huttner WB, and Kosodo Y (2005). Symmetric versus asymmetric cell division during neurogenesis in the developing vertebrate central nervous system. *Curr. Opin. Cell Biol* 17, 648–657. 10.1016/j.ceb.2005.10.005. [PubMed: 16243506]

57. Fu Y, Sander JD, Reyon D, Cascio VM, and Joung JK (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol* 32, 279–284. 10.1038/nbt.2808. [PubMed: 24463574]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- Cumulative lineage barcodes record developmental dynamics
- Phylotime infers time-scaled cell phylogenies from lineage barcodes
- ICE-FASE reconstructs progenitor hierarchy and dynamics from time-scaled phylogenies
- Progenitor State Coverage (PScov) measures the robustness of fate map inferences

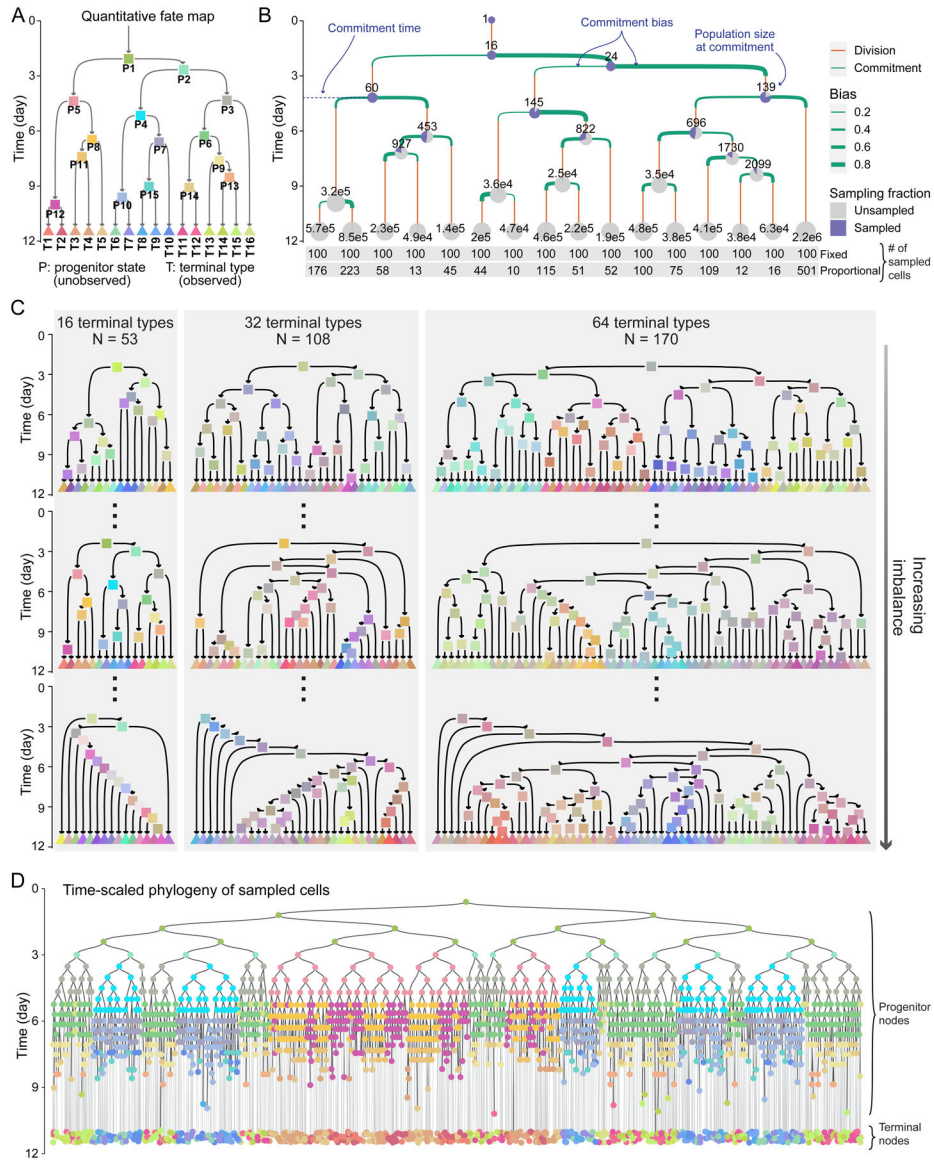


Figure 1. Simulating time-scaled phylogenetic trees of sampled cells based on a panel of quantitative fate maps.

(A) Topology of a quantitative fate map. Arrows represent cell states, colored rectangles represent their commitment events. Triangles represent terminal types at the time of sampling.

(B) Quantitative fate map in A annotated with commitment time, population size, and commitment bias of its progenitor states. Pie charts show sampling fraction under fixed sampling at progenitor state commitment time or terminal type sampling; numbers on pies show corresponding true population sizes. The bottom two rows show the number of cells sampled from each terminal type under fixed and proportional sampling.

(C) The panel of 331 quantitative fate maps. The maps are in three sizes of 16, 32, and 64 terminal cell types. Three examples of each size are shown, including the most balanced (top) and the most unbalanced (bottom).

(D) Example time-scaled phylogeny of sampled cells generated by fixed sampling of 100 cells from each terminal type in the fate map shown in **A**. Node colors based on **A**. See also Figures S1 and S2.

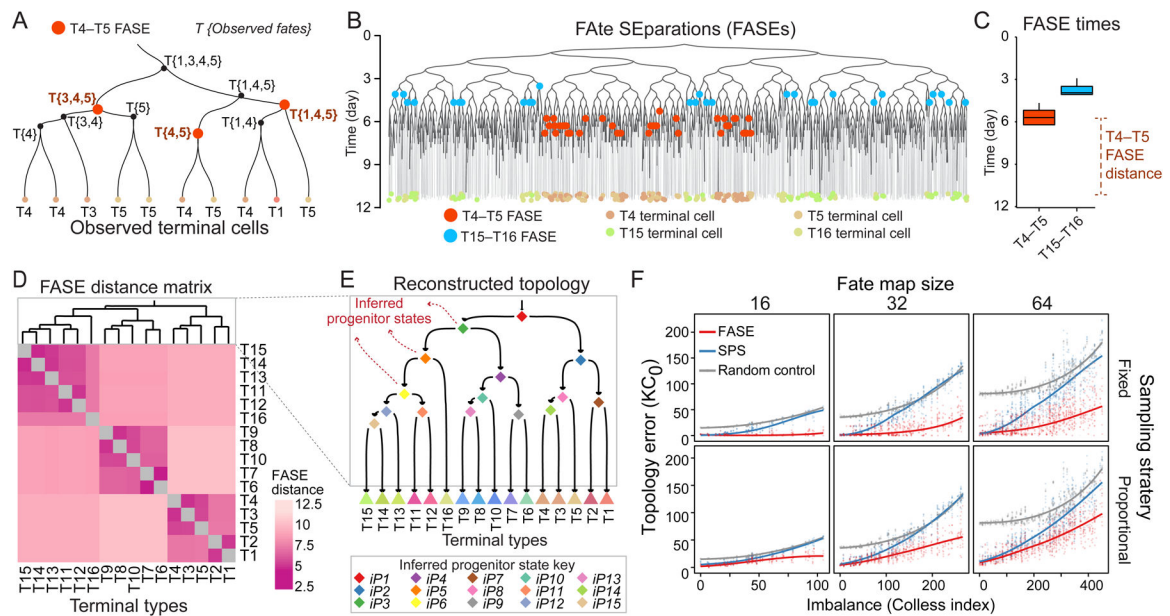


Figure 2. Reconstructing fate map topology from time-scaled phylogeny of sampled cells.

(A) Example phylogenetic subtree. Each internal node is labeled with its observed fate.

T4–T5 FASEs are colored orange.

(B) Phylogeny from Figure 1D showing T4–T5 (orange) and T15–T16 (blue) FASEs. T4, T5, T15, and T16 are colored according to the key on the bottom.

(C) Boxplots showing the temporal distribution of T4–T5 and T15–T16 FASEs in the tree in B.

(D) Heatmap showing the FASE distance matrix for all pairs of terminal types in the tree in B. Dendrogram shows hierarchical clustering result.

(E) The fate map topology reconstructed by clustering the FASE distance matrix in D.

Triangles: observed terminal types; diamonds: inferred progenitor states (*iPs*).

(F) Scatter plots showing error of fate map topology reconstruction (KC_0) using FASE algorithm (red) or SPS (blue) as a function of imbalance for all 3,310 simulated phylogenies faceted by fate map size (columns) and sampling strategy (rows). The gray points are based on random topology reconstructions. Solid trend lines are locally weighted smoothing (LOESS).

See also Figure S3.

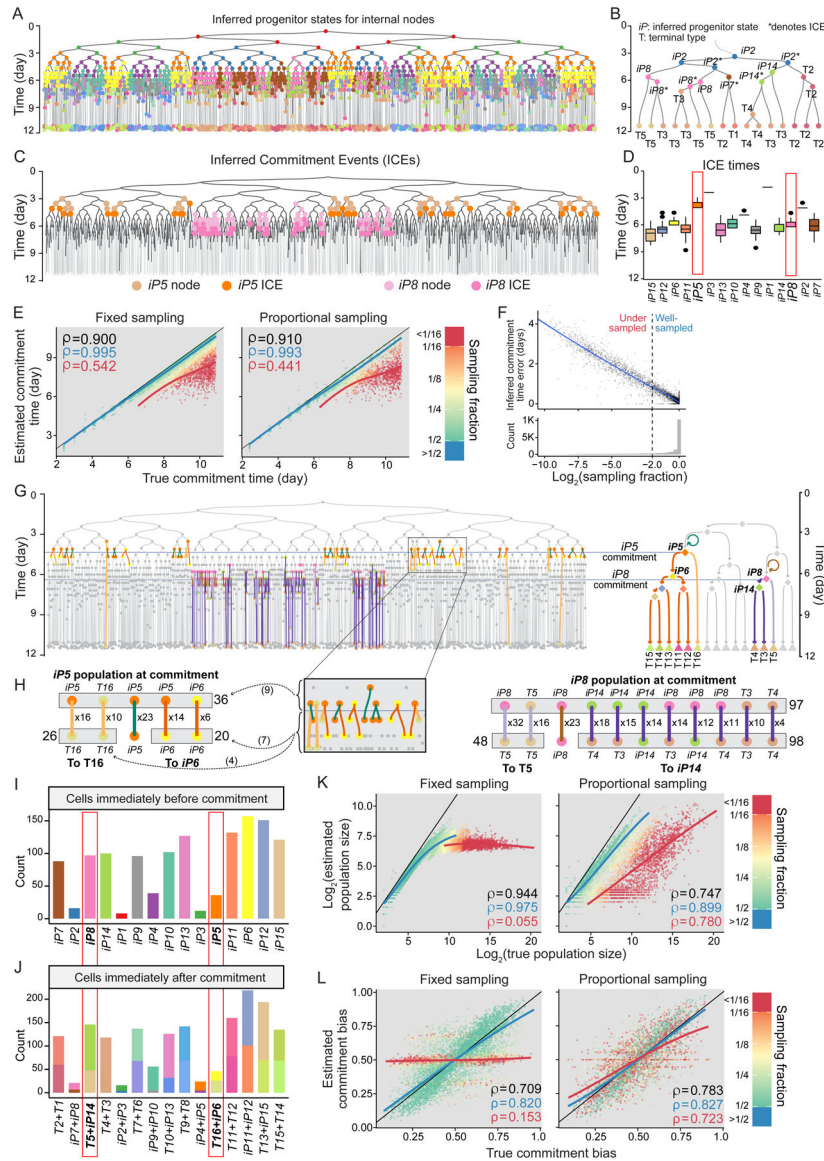


Figure 3. Obtaining progenitor state commitment times, population sizes, and commitment biases from time-scaled phylogeny of sampled cells.

(A) Phylogeny in Figure 2B with internal nodes colored according to their inferred progenitor state. Color key in Figure 2E.

(B) A subtree from A where each internal node is labeled by its inferred progenitor state or terminal type. Asterisks signify ICE nodes.

(C) The tree in A with nodes inferred as *iP5* (orange) and *iP8* (pink) labeled. Darker shades mark the nodes that are also ICEs.

(D) Boxplots showing the distribution of ICE times for all inferred progenitor states in the tree in A, representing their commitment times. The ICE times for *iP5* and *iP8* are boxed in red.

(E) Scatterplots showing the correlation between true commitment time of each progenitor state to the value estimated from the phylogenetic tree across all 3,310 simulated phylogenies broken down by sampling strategy. Dot colors represent progenitor states’

sampling fractions based on the key on the right. Blue line and value respectively show trendlines (LOESS) and Spearman's ρ for progenitors with sampling fraction $\geq 25\%$. Those for progenitor states with sampling fraction $< 25\%$ are shown in red. Spearman's ρ for all progenitor states is shown in black. Black lines are $y=x$.

(F) Scatter plot showing the error of inferred commitment time for progenitor states as a function of their sampling fraction (top) aligned to a histogram of sampling fraction for all progenitor states in the panel (bottom). Error is the absolute value of the difference between estimated and true commitment times. The vertical dashed line shows sampling fraction cutoff of 0.25. Trendline (LOESS) is in blue.

(G) The edges and the nodes relevant to population size and commitment bias of *iP5* and *iP8* are shown on the tree from **A** (left) aligned to the corresponding reconstructed fate map (right). Horizontal lines mark *iP5* and *iP8* commitment times. Edges in the phylogeny are classified and colored based on the commitment they represent, and nodes based on their inferred progenitor state or terminal type, with the fate map serving as the color key. Circular arrows on the fate map indicate divisions without commitment (self-renewal).

(H) Tally of edges and pre- and post-commitment nodes associated with *iP5* (left) and *iP8* (right) from the phylogeny in **G**, a part of which is magnified for added clarity. The number next to each edge indicates how many times the combination of incoming and outgoing progenitor state or terminal type was observed. Gray boxes on top indicate the nodes that were counted towards progenitor state population size, gray boxes on the bottom indicate post-commitment nodes counted for commitment bias. The number next to each gray box indicates the total count of nodes in that box. Arrows from **G** inset show in parenthesis counts being added to relevant tallies from the inset. Color key same as **G**.

(I) Barplots showing the estimated population size of each inferred progenitor state in **A**. Red boxes mark *iP5* and *iP8* estimates from **H**.

(J) Stacked barplots showing the estimated post-commitment population size of each inferred progenitor state in **A**, stacked and colored according to the downstream state they lead to. Red boxes mark estimates relevant to *iP5* and *iP8*, colored according to Figure 2E.

(K,L) Scatter plots showing the correlation between true and estimated population size (**K**) and commitment bias (**L**) of inferred progenitor states in fixed (left) and proportional sampling (right) across all 3,310 simulated phylogenies. Other plot features identical to panel **E**.

See also Figure S3.

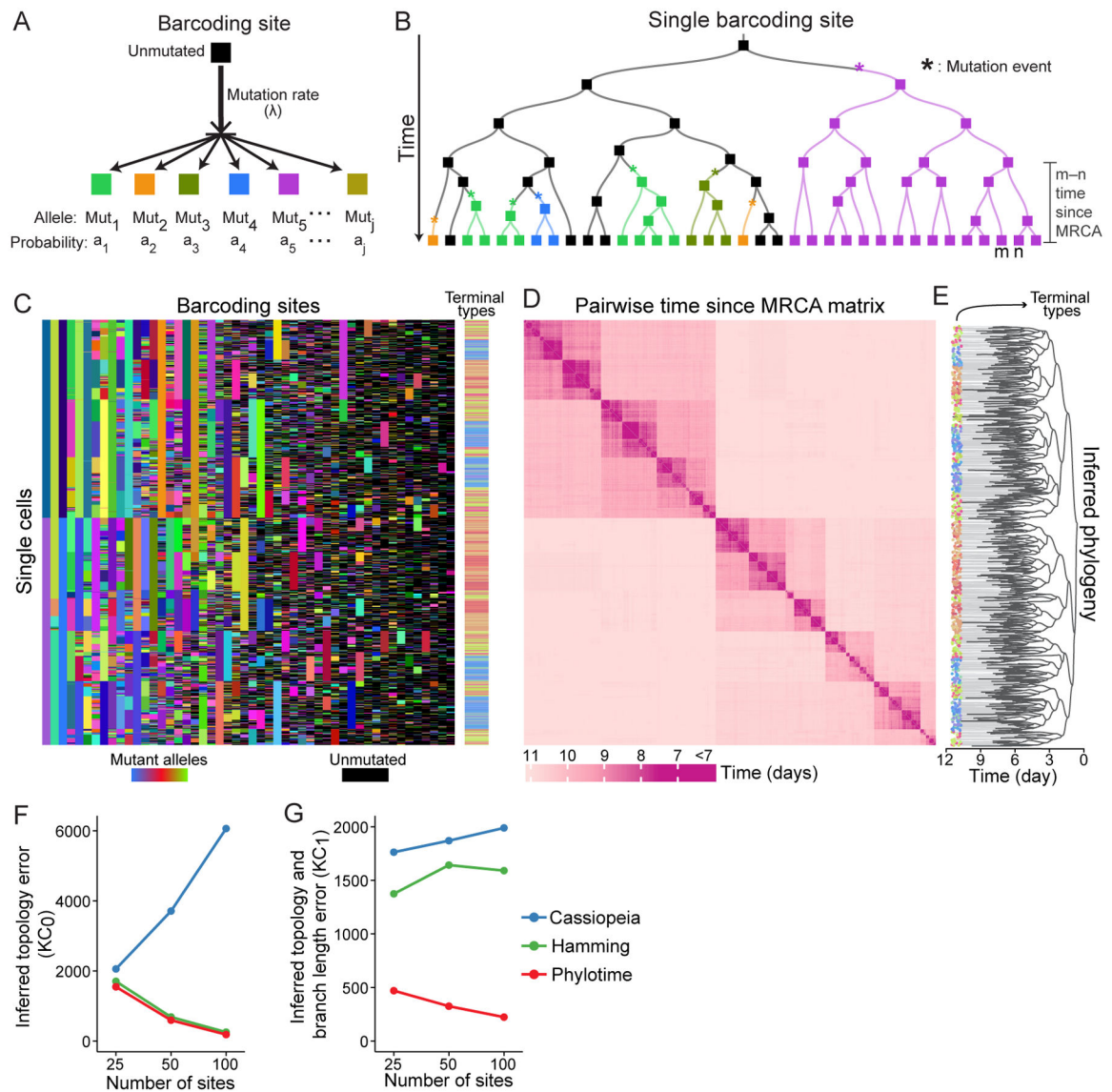


Figure 4. Modeling barcoding mutagenesis and inferring time-scaled phylogenies from lineage barcodes using Phylotime.

(A) Model of a barcoding site which mutates with a constant rate (λ) to convert to one of many possible mutant alleles (Mut_1, \dots, Mut_j), each with an emergence probability (a_1, \dots, a_j).

(B) Example time-scaled phylogeny showing the inheritance of a single barcoding site. Rectangles represent the site, their colors represent its allelic state according to A. Asterisks denote mutation events. Only the unmutated allele (black) can change, mutated alleles (colored) are inherited from a cell by all its descendants. The time since MRCA is shown for two terminal cells, m and n .

(C) Character matrix showing the output of a simulated barcoding experiment based on Figure 1A,B fate map with each barcoding site as a column and each cell as a row. Unmutated alleles are in black, mutant alleles in other colors. Color bar on the right shows cell types with Figure 1A as color key.

(D) Heatmap showing the estimated pairwise time since MRCA for all cells in **C**.
(E) Time-scaled phylogenetic tree inferred by applying Phylotime to **D**. Colors of terminal nodes signify cell type.
(F,G) Error of phylogenetic reconstruction using Phylotime (red), Hamming distance with UPGMA (green), and Cassiopeia (blue), with 25, 50, or 100 barcoding sites when considering only tree topology (**F**) or tree topology and branch length (**G**) across the panel of 530 simulated barcoding experiments with 16 terminal types (Mean \pm SEM, N=530; SEM very small).
See also Figure S4.

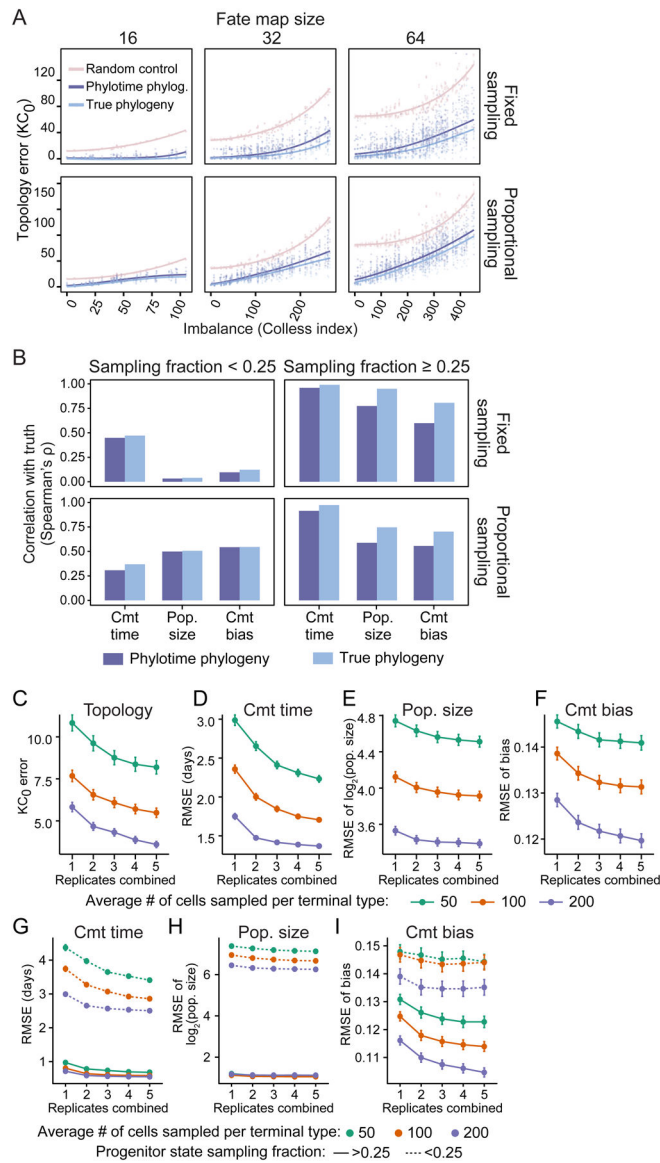


Figure 5. Quantitative fate map reconstruction from lineage barcodes using Phylotime and ICE-FASE.

(A) Scatter plots showing error of fate map topology reconstruction (KC_0) after applying ICE-FASE algorithm to true (light blue) or Phylotime-inferred (dark blue) phylogeny as a function of map imbalance for all 3,310 simulated barcoding experiments faceted by fate map size (columns) and sampling strategy (rows). The faded pink points establish baseline error based on random topology reconstruction. Trendlines (LOESS) are shown.

(B) Barplots of Spearman's correlation between true and estimated commitment (cmt) time, population (pop.) size, cmt bias for all progenitor states in 3,310 simulated experiments, faceted by sampling fraction (columns) and sampling strategy (rows). Values from true phylogenies are shown in light blue and those from Phylotime-inferred in dark blue.

(C–F) Lineplots showing the performance of quantitative fate map reconstruction as a function of the number of replicates combined when sampling 50 (green), 100 (orange), or 200 (purple) cells on average from each terminal type in the panel of 16-terminal type fate

maps, measured by the error of topology (**C**), cmt time (**D**), pop. size (**E**), and cmt bias (**F**) estimates (Mean \pm SEM; N=106).

(G–I) Same as **D–F** but estimates separated for well-sampled (solid lines) and undersampled (dotted lines) progenitor states.

See also Figures S4, S5, and S6.

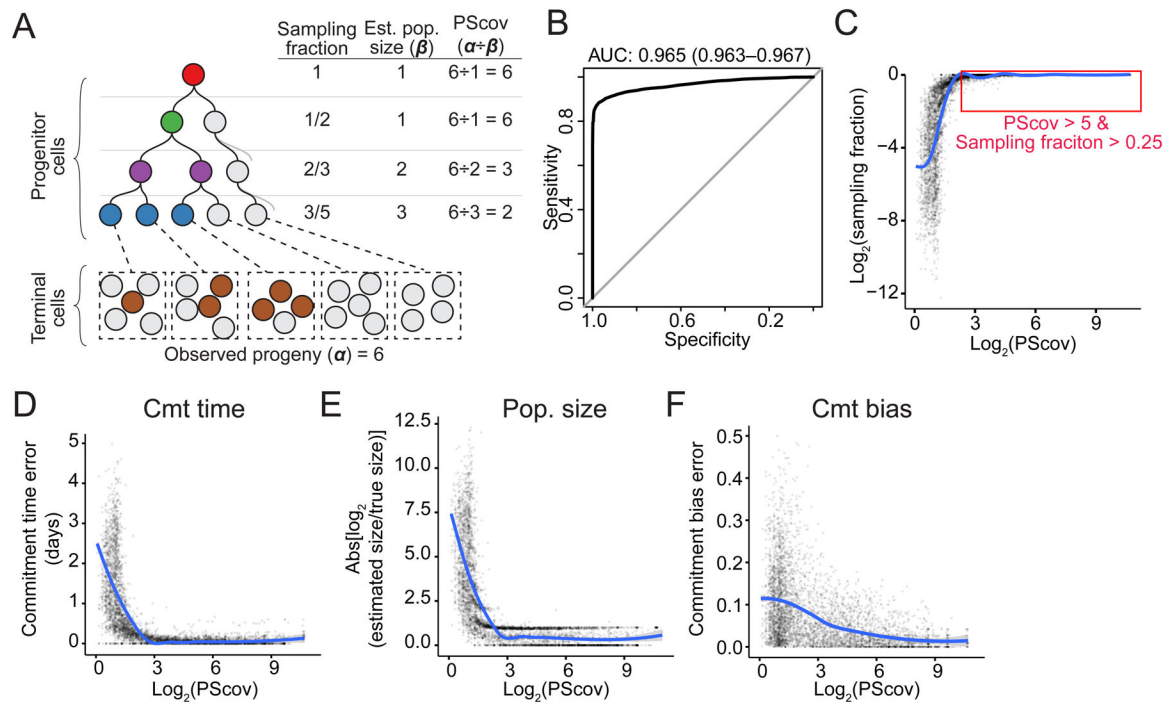


Figure 6. PScov measures robustness of reconstructed quantitative fate maps.

(A) Schematic showing how PScov is calculated. Each circle represents a cell, colors represent different progenitor states the cells have been assigned to, gray circles represent unsampled progenitor and terminal cells. Est: estimate; pop: population.

(B) Area Under the Receiver Operating Characteristics (auROC) curve for PScov shows a high sensitivity and specificity in detecting well-sampled progenitor states. CI: confidence interval.

(C) Scatter plot of progenitor state sampling fraction as a function of its PScov in 3,310 simulated experiments. Trendline (LOESS) shown in blue.

(D–E) Scatter plots showing the error of cmt time (D), pop. size (E), and cmt bias (F) estimates for progenitor states as a function of their PScov in our panel of 3,310 simulated experiments.

Trendline (LOESS) shown in blue.

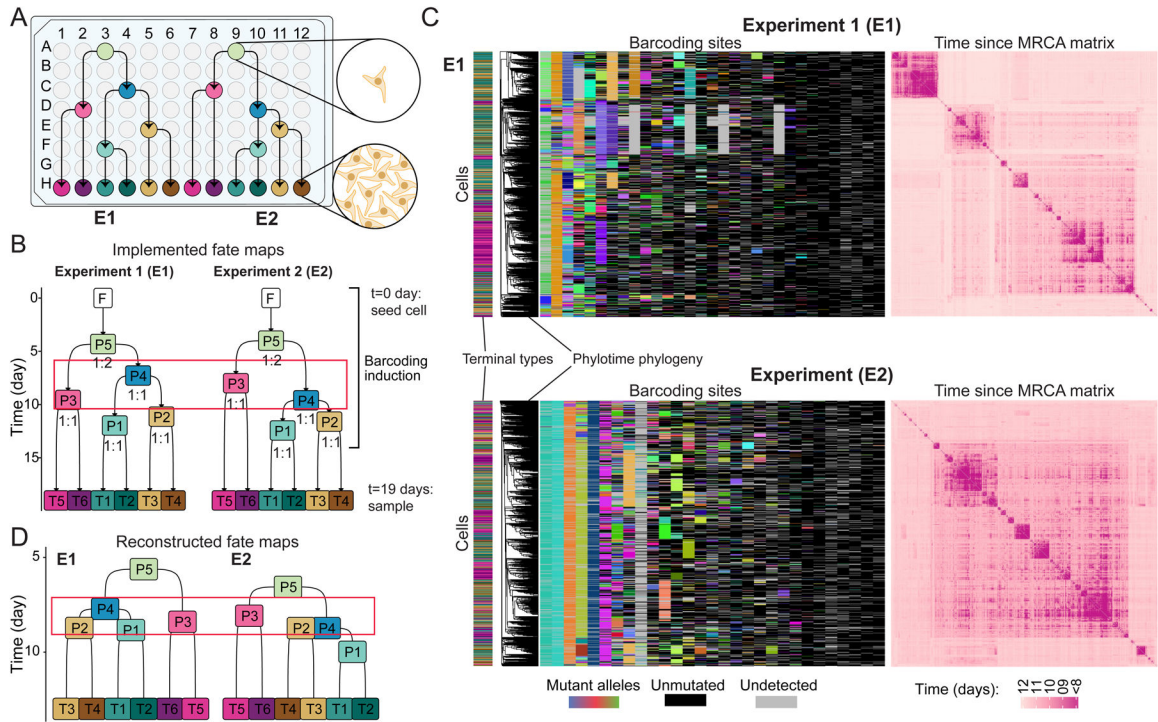


Figure 7. Validation of quantitative fate mapping using cell culture model system.
(A) Split-passage scheme in cultured iPSCs for E1 and E2 experiments. Arrows show passages.
(B) Quantitative fate maps implemented in E1 and E2 using scheme in A. Numbers under progenitor states mark split ratios (left:right); red box highlights the differences between progenitor state order. F: founder cell; P: progenitor states; T: terminal type.
(C) Character matrices of lineage barcodes (left) and heatmaps of pairwise time since MRCA matrix from hundreds of sequenced single cells in E1 (top) and E2 (bottom). Phylotime-inferred phylogeny is aligned to the matrix. Color bar on left marks the type of each cell according to B. Other plot features are the same as Figures 4C,D.
(D) Reconstructed quantitative fate maps for E1 (left) and E2 (right). The red box highlights the differences in progenitor state ordering. Labels same as B.
 See also Figure S7.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
AAVS1-Neo-M2rtTA	DeKolver et al. ⁵⁰	Addgene Cat#60843
Chemicals, peptides, and recombinant proteins		
Accutase™	STEMCELL Technologies	Cat#07920
Antibiotic-Antimycotic (100X)	ThermoFisher	Cat#15240062
Blasticidin S hydrochloride	MilliporeSigma	Cat#15205
(-)-Blebbistatin	MilliporeSigma	Cat#B0560
CloneR	STEMCELL Technologies	Cat#05888
Doxycycline hyclate	MilliporeSigma	Cat#D9891
DreamTaq Hot Start PCR Master Mix	ThermoFisher	Cat#K9012
Geneticin™ Selective Antibiotic (G418 Sulfate)	ThermoFisher	Cat#10131035
KAPA HiFi HotStart ReadyMix	Roche	Cat#07958935001
KAPA SYBR FAST qPCR Kits	Roche	Cat#07959389001
Lipofectamine™ Stem Transfection Reagent	ThermoFisher	Cat#STEM00001
Matrigel Growth Factor Reduced Basement Membrane Matrix	Corning	Cat#354230
mTeSR™ Plus	STEMCELL Technologies	Cat#100-0276
Opti-MEM I Reduced Serum Media	ThermoFisher	Cat#31985062
Pifithrin- α hydrobromide	Tocris Bioscience	Cat#1267
Puromycin dihydrochloride	MilliporeSigma	Cat#P8833
QuickExtract™ DNA Extraction Solution	Lucigen	Cat#QE09050
SYBR Green I Nucleic Acid Gel Stain	ThermoFisher	Cat#S7563
Critical commercial assays		
DNA Clean & Concentrator-5	Zymo Research	Cat#D4014
MiSeq Reagent Micro Kit v2 (300-cycles)	Illumina	Cat#MS-103-1002
MycoAlert® PLUS Mycoplasma Detection Kit	Lonza	Cat#LT07-701
Qubit dsDNA HS Assay Kit	ThermoFisher	Cat#Q32851
Deposited data		
Single-cell hgRNA sequencing raw FASTQ files for E1 and E2	This study	SRP386685
MARC1 sequencing data for determining hgRNA mutation rates	Kalhor et al. ⁵ Leeper et al. ³⁰	SRP155997
All quantitative fate maps	This study	10.5281/zenodo.7112097
inDelphi predicted mutant allele probabilities for hgRNAs in MARC1 mice and iPSC line	This study	10.5281/zenodo.7112097
Simulated phylogenies, sets of MARC1 hgRNAs used, single cell lineage barcodes, Phylotime reconstructed phylogenies for all experiments	This study	10.5281/zenodo.7112097
Experimental models: Cell lines		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
iPSC line: EP1-Cas9-hgRNA	This study	N/A
Oligonucleotides		
See Table S3 for oligonucleotides used in this study.	This study	N/A
Recombinant DNA		
Modified pSpCas9(BB)-2A-Puro (PX459) V2.0	Eldred et al. ⁴⁵	N/A
Modified Puro-Cas9 donor	Eldred et al. ⁴⁵	N/A
PB-U6insert hgRNA library	Kalhor et al. ⁵	Addgene Cat#104536
PB-U6insert-EF1puro library	Kalhor et al. ⁵	Addgene Cat#104537
Super piggyBac Transposase expression vector	System Biosciences	Cat#PB210PA-1
Software and algorithms		
Cassiopeia	Jones et al. ³⁵	https://github.com/YosefLab/Cassiopeia
ICE-FASE	This study	https://github.com/Kalhor-Lab/QFM/ DOI: 10.5281/zenodo.7114804
ImageJ	Schneider et al. ⁵¹	https://github.com/imagej/ImageJ
InDelphi	Shen et al. ³¹	https://github.com/rmaxwshen/inDelphi-model
Phylotime	This study	https://github.com/Kalhor-Lab/QFM/ DOI: 10.5281/zenodo.7114804
MARC1 analysis pipeline	Leeper et al. ³⁰	https://github.com/Kalhor-Lab/MARC1-Pipeline
Other		