Supplementary Materials and Methods

1. CLiNC algorithm

1.1 Calculating normalized covariances

Sample means and covariances are calculated from the table of counts indicating the number of cells with each barcode in each cell type. Let X_{ib} denote the number of cells with barcode b in cell type i. For each pair of cell types i and j, the normalized covariance can be calculated as

$$\text{Normalized covariance}(i,j) = \frac{1}{\mu_i \mu_j} \sum_{b=1}^{N} (X_{ib} - \mu_i)(x_{jb} - \mu_k) \quad \text{where} \;\; \mu_i = \frac{1}{N} \sum_{b=1}^{N} X_{ib}$$

1.2 Neighbor joining

Neighbor joining is performed iteratively. Each iteration begins with a matrix of barcode counts X_{ib}. Normalized covariance is calculated for all i, j pairs and the pair with the highest value is merged. During the merger, a new barcode counts matrix X' is constructed, in which the columns X_{i*} and X_{j*} are removed and a new column X'_{k*} representing the merger of i and j is created via $X'_{kb} = X_{ib} + X_{jb}$

1.3 Detecting symmetry violations

Let $\tilde{C}_{jk}$ denote the normalized covariance between nodes j and k. Symmetry violations are detected by bootstrapping over clones to resample the matrix of normalized covariances and then recording all the i, j, k triples that topologically satisfy conformal symmetry, yet

$$P_{\text{bootstrap}}\left(\tilde{C}_{jk} - \tilde{C}_{ik} > \epsilon\right) > FDR$$

where the threshold ϵ represents the maximum allowed deviation from perfect symmetry, and can in practice be set as $\epsilon = \text{median}(\tilde{C}_{jk} - \tilde{C}_{ik} \mid \text{for all putatively symmetric triples } i, j, k)$.

1.4 Inferring cross-tree transitions

Let $\mathcal{V}$ be the set of symmetry violations obtained in the previous step. For each potential cross tree transition $j' \to i'$ let $\mathcal{S}(j', i')$ be the set of violations that would be expected to occur if the transition existed (see Figure 3 and Appendix Theorem 6). Our goal is to parsimoniously explain the observed violations $\mathcal{V}$ as a union of predicted violations $\mathcal{S}(j', i')$. In doing so, we wish to cover as much of $\mathcal{V}$ as possible, while avoiding the inclusion of transitions for which only a minority of predicted violations belong to $\mathcal{V}$. This can be formalized as follows: Define a cost function

$$c(j', i') = 1 + \left(1 - \frac{|\mathcal{S}(j', i') \cap \mathcal{V}|}{|\mathcal{S}(j', i')|}\right)$$

and then find the set of transitions $j' \to i'$ that collectively cover $\mathcal{V}$ while minimizing the total cost $\sum c(j', i')$, and removing at the end any transitions with only a minority of violations in $\mathcal{V}$ (i.e. $c(j', i') < 1.5$). This optimization is equivalent to the well-known "set-cover problem", which is NP-complete but can be solved approximately. We use SetCoverPy (https://github.com/guangtunbenzhu/SetCoverPy) for approximate optimization.

1.5 Post-hoc correction of tree errors caused by cross-tree transitions

A post-hoc correction strategy is included based on simulation data (see section 7 of the Supplementary Methods). Distortions of the normalized covariance by cross-tree transitions tend to result in a characteristic pattern of transitions (identified in step 1.4 above) in which the transitions are chained ($i \rightarrow j$ and $j \rightarrow k$) and the parent of k is the sister of j. In such cases the inferred tree is usually incorrect and can be corrected by swapping the positions of j and the node that was assigned as a sister of *k*. The cross-tree transition ($i \rightarrow j$) is likely correct in these cases and should is retained. This situation is detected and corrected as a final step in the CLiNC pipeline.

2 *Simulations for success and failure cases (Figure 2)*

For the successful case (Figure 2A-C), a tree model was constructed as in Figure 2A (top). 200 single cells were initialized at the root node “0”. Following the distributions in Figure 2A (bottom), each cell generated cells at child nodes “1" and “4”, and cells at “1” in turn generated cells at nodes “2” and “3”. Normalized covariance was calculated for the resulting matrix of cell counts at leaf nodes, and neighbor joining was carried out as described in Results (“Recipe for data analysis”). For the failure case (Figure 2D-F), the same procedure was followed as above, but cell growth and partitioning were carried out according to the distributions in Figure 2D (bottom).

3 *Calculating the number of symmetry violations caused by random cross-tree transitions*

Random trees with 10 leaves were constructed as described below (Methods section “Robustness tests: Generating random trees and barcode distributions”). For each tree, the set of all symmetries was found, and then predicted symmetry violations were calculated for random cross-tree transitions, as specified by Theorem 6 and Figure 3. The proportion of violations (out of the total set of symmetries) was recorded. This procedure was performed for 2000 random trees and 10 cross-tree transitions per tree.

4 *Robustness tests: Generating random trees and barcode distributions*

Trees were generated using an inhomogeneous branching process. Beginning with a single root node, each node was either assigned to be a leaf (termination of branching) or an internal node with two children (continuation of branching). The probability of termination rose with increasing distance from the root node: $P(\text{termination}) = 1 - 0.7^{d-1}$ where d is distance to the root. Differentiation was simulated independently for each ‘barcode’ by initializing a single cell at the root node, and then at each stage assigning to each cell a number of children sampled from a Poisson distribution with mean 3, and partitioning the cells binomially to daughters in equal proportions.

5 *Robustness tests: Simulation of self-renewal*

Self-renewal simulations were carried out on trees with 10 leaves, generated as described above using 5000 barcode clones. Differentiation was again simulated independently for each ‘barcode’, initializing a single cell at the root node. At each time step each cell was assigned a number of children sampled from a Poisson distribution with mean 3 and these cells were partitioned between the two daughter nodes and the parent. Partitioning was performed with equal proportions at the root node (1/3, 1/3, 1/3) and in proportions that were biased toward the daughter cells for non-root nodes (2/5, 2/5, 1/5). At leaf nodes cells were either entirely replaced by incoming cells from the parent (“turnover model”), retained without further expansion (“accumulation model”) or allowed to remain and continue expanding (“expansion model”; burst size ~ Poisson(3)).

6 *Robustness tests: Cross-tree transitions*

Cross-tree transitions were simulated on trees with 10 leaves and 5000 barcodes. Transitions ($j' \rightarrow i'$) were randomly generated according to the following criteria: (1) j' is one step above i' in the tree, i.e. one step closer to the root; (2) j' is not the parent of i'; (3) No node can participate in more than one transition.

For a given tree and set of transitions, differentiation was carried out as described above (see "Robustness tests: Generating random trees and barcode distributions") with the following change: at the source nodes of transitions, cells were partitioned among the two canonical daughters and the transition target in proportions $((1-P/3)/2,\ (1-P/3)/2,\ P/3)$ where P denotes the "cross-tree transition probability". Symmetry violations were detected using a false-discovery rate of 1% and cross-tree transitions were detected using the CLiNC pipeline.

7 Robustness tests: Cross-tree transition tree correction

Cross-tree transitions frequently caused incorrect tree inference (Supp Figure 1H-I). We identified a characteristic pattern in these errors and a standard edit that could be used to recover the correct tree. The characteristic pattern is a pair of chained transitions ($i \rightarrow j$ and $j \rightarrow k$) where the parent of k is the sister of j. In such cases the inferred tree is usually incorrect and should be edited by swapping the positions of j and the node that was assigned as a sister of *k.* The cross-tree transition ($i \rightarrow j$) is likely correct and should be retained. This post-hoc correction is included in the CLiNC pipeline (see section 1.5 of the Supplementary Methods).

8 Robustness tests: Evaluation of accuracy

Accuracy of tree inference across all simulation is measured in two ways: "proportion correct" and "tree distance". Proportion correct refers to the fraction of cases where the inferred tree is identical the ground-truth tree. Since internal (non-leaf) nodes do not have intrinsic labels, two trees are considered identical of one obtains the same collection of leaf-sets when enumerating all subtrees. Tree-distance was measured using the Robinson-Foulds metric. We note that a Robinson-Foulds tree distance of zero does not imply that two trees are identical since they can still differ in the position of the root node.

9 Analysis of barcoding data in hematopoiesis

Barcoding data from a recent paper (1) were used (data available at GEO accession GSE140802). The paper includes data from a pilot transplantation study using a more mature starting population, and then a larger transplant study using an immature starting population. We restricted analysis to the latter dataset to better respect the modeling assumption of a uniform starting population. Cells were classified into cell types according to the clustering presented in the previous paper. Clusters from the same lineage at different stages of maturity were combined. Following the recommendations in Supplementary Figure 3A, we first removed cell types that are uncommitted progenitors of other cell types also measured in the experiment, including multipotent progenitors and granulocyte-monocyte progenitors. We then excluded rare cell types, defined as those with fewer than 200 shared barcodes (i.e. barcodes appearing in more than one cell type), including T cells, non-classical monocytes, megakaryocytes, eosinophils, macrophages and Ccr7+ dendritic cells. Cells from one-week post-transplant and two-weeks post-transplant were combined for analysis, under the rationale that different cell types have different time scales of maturation after commitment. For example, B cells were abundant at two-weeks post-transplant but almost absent one-week post-transplant. Tree construction and detection of symmetry violations were carried out as described in the Results. For bootstrap estimates of normalized covariance, we resampled clones with replacement 5000 times. Our analysis is fully reproducible with code available online (https://github.com/AllonKleinLab/CLiNC/blob/master/clinc_python/example/clinc_pipeline.ipynb).

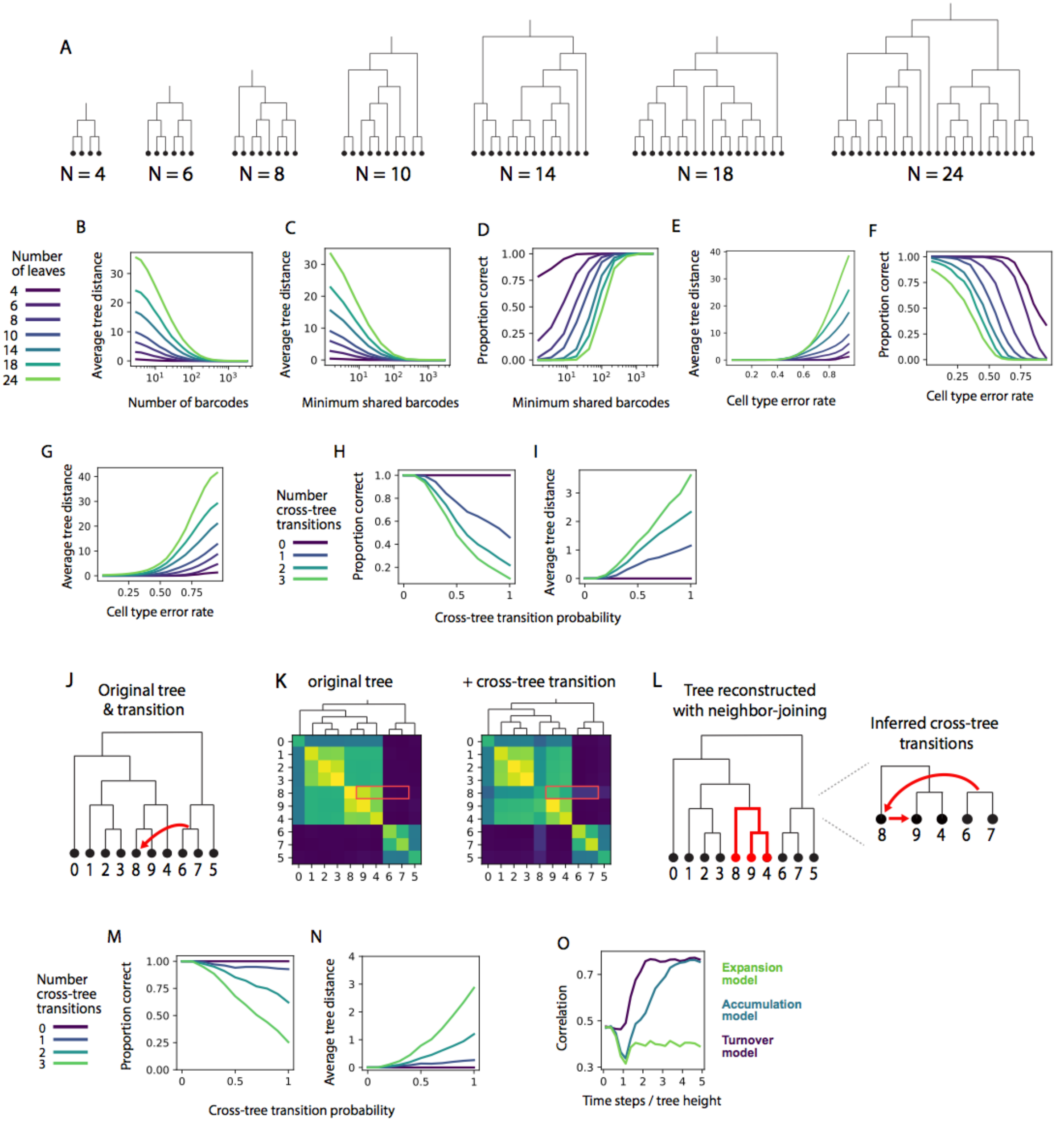
10 Analysis of cell cycle status in hematopoiesis

An aggregate cell cycle score was computed for each cell by Z-scoring and averaging expression of cell-cycle associated genes. The set of genes (N=85) was obtained from (2) and represents the top 20 Cyclebase (3) genes for each stage of the cell cycle. Only cell cycle scores for barcoded cells where used for Supp Figure 2D-E.

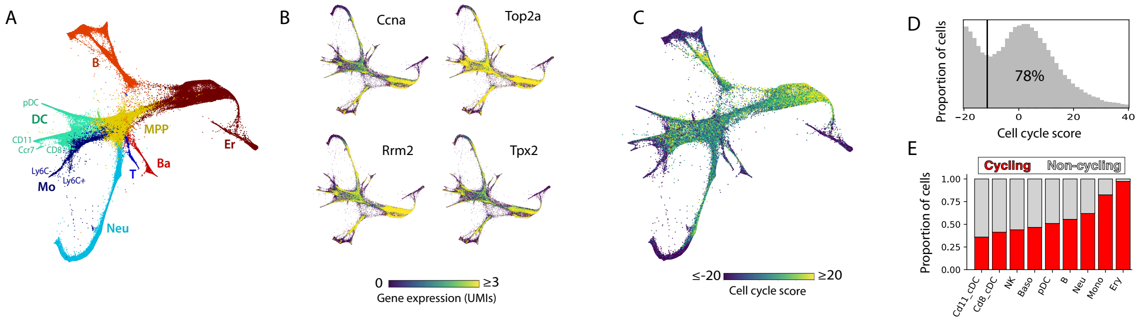
Supplemental Methods References

1. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, & Klein AM (2020) Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* 367(6479):eaaw3381.
2. Liu Z*, et al.* (2017) Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nat Commun* 8(1):22.
3. Santos A, Wernersson R, & Jensen LJ (2014) Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Research* 43(D1):D1140-D1144.

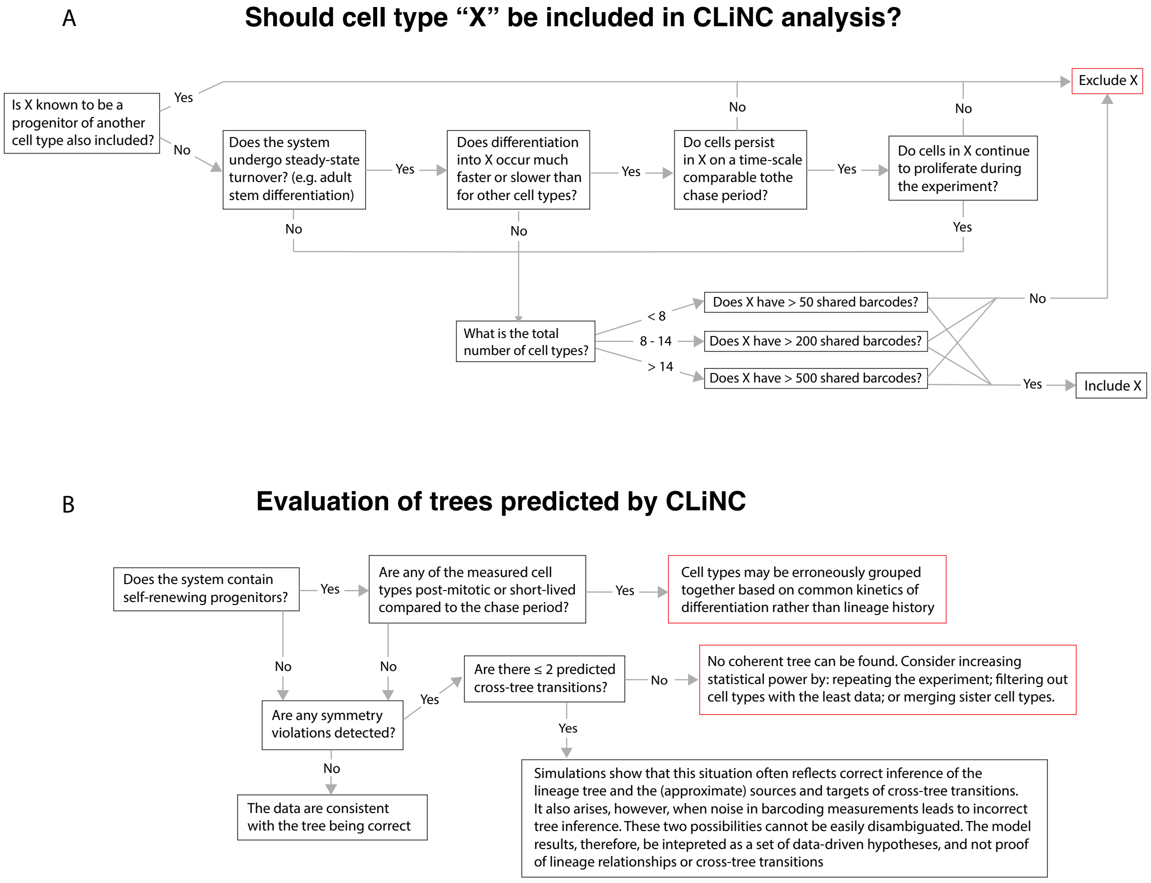
Supplementary Figures



Supplementary Figure 1. Robustness tests on simulated data (continued). (A) Examples of trees with varying numbers of leaves. (B) Same as Figure 4A, but showing average tree distance instead of proportion correct. (C-D) Same as Figure 4A and panel **B,** but the x-axis represents the minimum number of shared barcodes in any lineage, as opposed to the total number of barcodes in the simulation, where shared barcodes are defined as barcodes appearing in more than one lineage. (E) Same as Figure 4B, but showing average tree distance instead of proportion correct. (F-G) Same as Figure 4B and panel **E**, but with 500 barcodes used in the simulation rather than 5000. (H-J) Proportion of correctly inferred trees (H) or average tree distance (I) as a function of off-tree transition probability, in simulated trees with 0, 1, 2 or 3 cross-tree transitions, without any post-hoc correction. (J-L) Example showing the characteristic tree inference error caused by cross-tree transitions. (J) Differentiation is simulated in a tree with a transition from the parent of nodes "6” and “7” to node “8”. (K) Comparison of normalized covariance between a simulation without the transition (left) and with the transition (right; cross-tree transition probability = 0.5) shows an increased covariance between “8” and “5”,”6” and “7”, and a decreased covariance between “8” and all the members of its clade, especially “9” and “4”. (L) The tree inferred from normalized covariance erroneously places “8” in a separate branch upstream of “9” and “4”. CLiNC detects two cross-tree transitions, one from parent of “6” and “7” to “8” – reflecting the simulated cross-tree transition – and a second transition from “8” to “9”, reflecting the fact that they are sisters in the ground-truth tree and hence break symmetry with node “4”. Swapping nodes “8” and “4” in the inferred tree corrects the error and makes it identical to the ground truth tree. The pattern in this example general and can be used to systematically correct errors. (M-N), same as **H-I**, but with the corrections applied as described in **L.** (O) Average correlation between [difference in ground-truth distance to the root] and [distance to each other in inferred tree], for all pairs of leaf nodes, with separate averages taken for each timepoint and model of behavior at leaf nodes (see Figure 4E-F) for definition of models. Rising correlation in the turnover and accumulation models indicates that cell types are increasingly grouped together by timing of differentiation (distance from root) rather than shared lineage history.



Supplementary Figure 2. Analysis of cell cycle status in hematopoietic data. (A) SPRING plot reproduced from (14) showing cells from the hematopoietic barcoding dataset arranged using a force-directed graph layout. (B) Expression of cell cycle marker genes on the SPRING plot. (C-D) A cell cycle score was computed by averaging the Z-scored expression of 85 cell cycle-associated genes. (C) Plot of cell cycle score on the SPRING layout. (D) Histogram of cell cycle scores with threshold marking the score above which cell are considered to be cycling. (E) Proportion of cycling cells in each lineage.



A Should cell type “X” be included in CLiNC analysis?

A
Is X known to be a progenitor of another cell type also included?
Yes
No
Exclude X
Does the system undergo steady-state turnover? (e.g. adult stem differentiation)
Yes
Does differentiation into X occur much faster or slower than for other cell types?
Yes
Do cells persist in X on a time-scale comparable tothe chase period?
Yes
Do cells in X continue to proliferate during the experiment?
No
No
No
No
Yes
What is the total number of cell types?
< 8
8 - 14
> 14
Does X have > 50 shared barcodes?
Does X have > 200 shared barcodes?
Does X have > 500 shared barcodes?
No
Yes
Include X

B Evaluation of trees predicted by CLiNC

B
Does the system contain self-renewing progenitors?
Yes
Are any of the measured cell types post-mitotic or short-lived compared to the chase period?
Yes
Cell types may be erroneously grouped together based on common kinetics of differentiation rather than lineage history
No
No
Are any symmetry violations detected?
Yes
Are there ≤ 2 predicted cross-tree transitions?
No
No coherent tree can be found. Consider increasing statistical power by: repeating the experiment; filtering out cell types with the least data; or merging sister cell types.
Yes
Simulations show that this situation often reflects correct inference of the lineage tree and the (approximate) sources and targets of cross-tree transitions. It also arises, however, when noise in barcoding measurements leads to incorrect tree inference. These two possibilities cannot be easily disambiguated. The model results, therefore, be intepreted as a set of data-driven hypotheses, and not proof of lineage relationships or cross-tree transitions
No
The data are consistent with the tree being correct

Supplementary Figure 3. Guidance data preprocessing decisions and tree interpretation. (A) Flow chart for deciding whether to include a cell type in the analysis. The criteria consider whether the differentiation kinetics, if they include self-renewal, are similar to the “Expansion model” (Figure 4) and whether the cell type was sampled densely enough for accurate inference. (B) Flow chart for interpretation of CLiNC output. If there is self-renewal, the criteria again assess similarity to the “Expansion model” – this time on a global basis. If symmetry violations are detected, the criteria assess whether they are consistent with the existence of cross-tree transitions and whether the structure of those transitions implies a possible error in reconstruction of the tree. All criteria across both **A** and **B** are derived from the simulations shown in Figure 4 and Supplementary Figure 1.

Theory Appendix

1 Branching process model

To model the propagation of barcoded cells during differentiation, we will use a multi-type branching process. Let $T = \{0, 1, ..., m\}$ be a rooted tree with integer-labeled nodes. Assume the root node is always called 0. The nodes of T represent cell types at various stages in development (assume N stages). Define a parent function $p : T \to T$ that maps each node to its parent. We will use p^n to denote the n-wise composition of p, so that for example $p^2(i) = p(p(i))$ and also for notational convenience, $p^0(i) = i$. As a barcoded clone divides and differentiates, it propagates down T, with different numbers of cells being deposited into each node $i \in T$. Let the random variable X_i represent the number of cells from a barcoded clone that enter node $i \in T$.

At each internal node of T, cells divide and differentiate. We will assume that the process of division and differentiation is independent and identical for each cell at a given node, though the process may differ between different nodes. Thus, for each single cell at an internal node $i \in T$, there is a distribution $D_i = P(\{X_j\}_{p(j)=i} \mid X_i = 1)$ over the number of new cells that will be passed to the child nodes of i. For a single cell beginning at the root of T, repeated action of the D_i distributions sends a cascade of daughter cells down the layers of T, producing at the end a collection of counts $\{X_{i_1}, X_{i_2}, ...\}$ at the leaf nodes. Here, we investigate whether the moments of these counts can be used to reconstruct T.

2 Probability generating functions

Probability generating functions will be our main tool for calculating the moments of the $\{X_i\}$. We review their definition and some key properties below.

Definition 1. *Let $X = (X_1, ..., X_m)$ be a multivariate random variable. The probability generating function (p.g.f.) of X is defined by*

$$\psi_X(z_1, ..., z_m) = \mathbb{E}(z_1^{X_1} z_2^{X_2} \dots z_m^{X_m}) \tag{1}$$

Property 1. *Let $X = (X_1, ..., X_m)$ be a multivariate random variable. The first and second moments of X can be calculated from ψ_X using the facts below, where $\mathbf{1}$ denotes the tuple $(1..., 1)$.*

$$\mathbb{E}(X_i) = \frac{\partial \psi_X}{\partial z_i}\bigg|_{\mathbf{1}} \qquad \mathbb{E}(X_i^2) = \left(\frac{\partial \psi_X^2}{\partial z_i^2} + \frac{\partial \psi_X}{\partial z_i}\right)\bigg|_{\mathbf{1}} \qquad \mathbb{E}(X_i X_j) = \frac{\partial^2 \psi_X}{\partial z_i \partial z_j}\bigg|_{\mathbf{1}} \tag{2}$$

Property 2. *Let $X = (X_1, ..., X_m)$ be a multivariate random variable, and let $Y = \{Y_1, ..., Y_m\}$ be a collection of independent random variables. Define a new multivariate random variable Z*

$$Z = \left(\sum_{i=1}^{X_1} (Y_1)_i,\ \sum_{i=1}^{X_2} (Y_2)_i,\ ...,\ \sum_{i=1}^{X_m} (Y_m)_i \right) \tag{3}$$

Where $(Y_1)_i$ is a sample from the variable Y_1. The probability generating function of Z is

$$\psi_Z(z_1, ..., z_m) = \psi_X(\psi_{Y_1}(z_1), ..., \psi_{Y_m}(z_m)) \tag{4}$$

Using Property 2, we can construct the p.g.f. of the branching process from Section 1 by chaining together the p.g.f.'s of each division and differentiation step. For each $i \in T$, let ψ_i denote the p.g.f. of the local division and differentiation distribution D_i. Note that the stochastic process defined in Section 1, where a single cell beginning at the root of T divides and differentiates down the tree, can be restricted to any subtree of T. Let Ψ_i be the p.g.f. of the process restricted to the subtree T_i rooted at i. Recalling our convention to name the root of T as 0, it follows that Ψ_0 is the p.g.f of the full process from Section 1. Ψ_0 can be calculated from the following recursion

$$\Psi_i = \psi_i(\Psi_{j_1} ..., \Psi_{j_k}) \quad \text{where} \quad j_1 ..., j_k \quad \text{are the child nodes of } \ i. \tag{5}$$

Note, we will generally use z to refer to the arguments of a p.g.f. When a p.g.f. describes a distribution of random variables $X_{i_1}, ..., X_{i_k}$, then each of its arguments corresponds to one of these random variables. Throughout the following text, the z arguments will be identified with their corresponding random variable by subscript, so that for example, z_i corresponds to X_i - which (as noted previously) denotes the number of cells at tree node i.

3 Calculation of moments

To reconstruct the topology of T, we will use the normalized covariances $\frac{\text{Cov}(X_i,X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)}$ for each pair of leaf nodes $i, j \in T$. These moments depend on the division and differentiation processes D_i at each internal node. To describe this relationship, it is useful to introduce the notation:

$$E_i := \mathbb{E}(X_i \mid X_{p(i)} = 1) \quad \text{and} \quad C_{i,j} := \text{Cov}(X_i, X_j \mid X_{p(i)} = 1) \ \text{ where } \ p(i) = p(j) \tag{6}$$

Theorem 1. *Let i be a leaf-node of T. Then*

$$\mathbb{E}(X_i) = \prod_{k=0}^{N-1} E_{p^k(i)} \ \ \textit{which follows from} \quad \frac{\partial \Psi_0}{\partial z_i} = \prod_{k=0}^{N-1} \frac{\partial \psi_{p^{k+1}(i)}}{\partial z_{p^k(i)}} \tag{7}$$

Proof. Apply the chain rule to the recursion in line (4). □

Theorem 2. *Let i, j be leaf-nodes of T and M the smallest integer with $p^M(i) = p^M(j)$. Such an M exists because T is rooted and there are always N steps from root to leaf. Our key result is*

$$\frac{\text{Cov}(X_i, X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} = \sum_{m=M-1}^{N-1} \frac{1}{\mathbb{E}(X_{p^{m+1}(i)})} \left(\frac{C_{p^m(i),p^m(j)}}{E_{p^m(i)} E_{p^m(j)}} \right) \tag{8}$$

Proof. Note that $\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)$, which implies

$$\frac{\text{Cov}(X_i, X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} = \frac{1}{\mathbb{E}(X_i)\mathbb{E}(X_j)} \left(\frac{\partial^2 \Psi_0}{\partial z_i \partial z_j} \right) \bigg|_{\mathbf{1}} - 1 \tag{9}$$

To compute $\partial^2\Psi_0/(\partial z_i \partial z_j)$, note that

$$\frac{\partial^2\Psi_0}{\partial z_i \partial z_j} = \sum_{m=0}^{N-1}\left(\prod_{k=0}^{m-1}\frac{\partial\psi_{p^{k+1}(i)}}{\partial z_{p^k(i)}}\right)\frac{\partial}{\partial z_j}\left(\frac{\partial\psi_{p^{m+1}(i)}}{\partial z_{p^m(i)}}\right)\left(\prod_{k=m+1}^{N-1}\frac{\partial\psi_{p^{k+1}(i)}}{\partial z_{p^k(i)}}\right) \tag{10}$$

$$\text{and} \quad \frac{\partial}{\partial z_j}\left(\frac{\partial\psi_{p^{m+1}(i)}}{\partial z_{p^m(i)}}\right) = \left(\prod_{k=0}^{m-1}\frac{\partial\psi_{p^{k+1}(j)}}{\partial z_{p^k(j)}}\right)\left(\frac{\partial\psi^2_{p^{m+1}(i)}}{\partial z_{p^m(i)}\partial z_{p^m(j)}}\right) \tag{11}$$

where line (10) follows from the product rule and line (11) follows from the chain rule. The empty products $\prod_{k=N}^{N-1}$ and $\prod_{k=0}^{-1}$ are always equal to 1 and appear in in line (10) above out of notational convenience. Together lines (10) and (11) imply (12) below, where terms in the sum with $m < M-1$ are ommitted due to vanishing of the second differential.

$$\frac{\partial^2\Psi_0}{\partial z_i \partial z_j} = \sum_{m=M-1}^{N-1}\left(\prod_{k=0}^{m-1}\frac{\partial\psi_{p^{k+1}(i)}}{\partial z_{p^k(i)}}\frac{\partial\psi_{p^{k+1}(j)}}{\partial z_{p^k(j)}}\right)\left(\frac{\partial\psi^2_{p^{m+1}(i)}}{\partial z_{p^m(i)}\partial z_{p^m(j)}}\right)\left(\prod_{k=m+1}^{N-1}\frac{\partial\psi_{p^{k+1}(i)}}{\partial z_{p^k(i)}}\right) \tag{12}$$

Applying Theorem 1 now gives

$$\frac{1}{\mathbb{E}(X_i)\mathbb{E}(X_j)}\left(\frac{\partial^2\Psi_0}{\partial z_i \partial z_j}\right) = \sum_{m=M-1}^{N-1}\frac{1}{E_{p^m(i)}E_{p^m(j)}}\left(\frac{\partial\psi^2_{p^{m+1}(i)}}{\partial z_{p^m(i)}\partial z_{p^m(j)}}\right)\left(\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(j)}}\right) \tag{13}$$

Using line (2), we obtain

$$\left.\frac{\partial\psi^2_{p^{m+1}(i)}}{\partial z_{p^m(i)}\partial z_{p^m(j)}}\right|_{\mathbf{1}} = \mathbb{E}(X_{p^m(i)}X_{p^m(j)}|X_{p^{m+1}(i)} = 1) - \delta_{p^m(i),p^m(j)}E_{p^m(i)} \tag{14}$$

$$= C_{p^m(i),p^m(j)} + E_{p^m(i)}E_{p^m(j)} - \delta_{p^m(i),p^m(j)}E_{p^m(i)} \tag{15}$$

where δ refers to the Kronecker delta. It follows that

$$\frac{1}{\mathbb{E}(X_i)\mathbb{E}(X_j)}\left.\left(\frac{\partial^2\Psi_0}{\partial z_i \partial z_j}\right)\right|_{\mathbf{1}} = \sum_{m=M-1}^{N-1}\left(\frac{C_{p^m(i),p^m(j)}}{E_{p^m(i)}E_{p^m(j)}} + 1 - \frac{\delta_{p^m(i),p^m(j)}}{E_{p^m(j)}}\right)\left(\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(i)}}\right) \tag{16}$$

We may now cancel some terms. Observe that

$$\sum_{m=M-1}^{N-1}\left(1 - \frac{\delta_{p^m(i),p^m(j)}}{E_{p^m(j)}}\right)\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(i)}} = \sum_{m=M-1}^{N-1}\left(\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(i)}} - \prod_{k=m}^{N-1}\frac{\delta_{p^m(i),p^m(j)}}{E_{p^k(i)}}\right) \tag{17}$$

$$= \sum_{m=M-1}^{N-1}\left(\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(i)}}\right) - \sum_{m=M-1}^{N-2}\left(\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(i)}}\right) = \prod_{k=N}^{N-1}\frac{1}{E_{p^k(i)}} = (\text{empty product}) = 1 \tag{18}$$

We may now simplify line (16) to show that

$$\frac{1}{\mathbb{E}(X_i)\mathbb{E}(X_j)}\left.\left(\frac{\partial^2\Psi_0}{\partial z_i \partial z_j}\right)\right|_{\mathbf{1}} = \sum_{m=M-1}^{N-1}\left(\frac{C_{p^m(i),p^m(j)}}{E_{p^m(i)}E_{p^m(j)}}\right)\left(\prod_{k=m+1}^{N-1}\frac{1}{E_{p^k(i)}}\right) + 1 \tag{19}$$

The main result can now be seen from

$$\frac{1}{\mathbb{E}(X_{p^{m+1}(i)})} = \prod_{k=m+1}^{N-1} \frac{1}{E_{p^k(i)}} \quad \text{and} \quad \frac{\mathrm{Cov}(X_i, X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} = \frac{1}{\mathbb{E}(X_i)\mathbb{E}(X_j)} \left(\frac{\partial^2 \Psi_0}{\partial z_i \partial z_j} \right)\bigg|_{\mathbf{1}} - 1 \tag{20}$$

□

4 Tree reconstruction

The normalized covariances calculated in Section 3 measure the extent to which barcodes jointly appear in any pair of cell types. It is appealing to reconstruct the topology T using a greedy approach where pairs of nodes having the highest normalized covariance are iteratively joined together (similar to Neighbor Joining approached in phylogenetic inference). The following theorem shows when this approach will work.

Theorem 3. *Let E_i and $C_{i,j}$ be the moments of the differentiation and division processes D_i, as defined in line (6). Suppose that for all pairs of sister nodes $a, b \in T$,*

$$\frac{C_{a,a}}{E_a^2} \geq \frac{C_{a,b}}{E_a E_b} + \frac{1}{E_a}, \quad \textit{and} \quad \frac{C_{a,b}}{E_a E_b} > -1 \tag{21}$$

And for all triplets of sister nodes $a, b, c \in T$,

$$\frac{C_{a,b}}{E_a E_b} = \frac{C_{a,c}}{E_a E_c} \tag{22}$$

Then for any leaves $i, j, k \in T$, i and j are more closely related than i and k if an only if

$$\frac{\mathrm{Cov}(X_i, X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} > \frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} \tag{23}$$

where "more closely related" means that there exists an integer M with $p^M(i) = p^M(j) \neq p^M(k)$.

Proof. Suppose that i and j and more closely related than i and k. There exist minimal M_1 and M_2 with $p^{M_1}(i) = p^{M_1}(j)$ and $p^{M_2}(i) = p^{M_2}(k)$ and they satisfy $M_1 < M_2$. From Theorem 2, we know

$$\frac{\mathrm{Cov}(X_i, X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} - \frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} = \frac{1}{\mathbb{E}(X_{p^{M_2}(i)})} \left(\frac{C_{p^{M_2-1}(i), p^{M_2-1}(j)}}{E_{p^{M_2-1}(i)} E_{p^{M_2-1}(j)}} - \frac{C_{p^{M_2-1}(i), p^{M_2-1}(k)}}{E_{p^{M_2-1}(i)} E_{p^{M_2-1}(k)}} \right) \tag{24}$$

$$+ \sum_{m=M_1-1}^{M_2-2} \frac{1}{\mathbb{E}(X_{p^{m+1}(i)})} \left(\frac{C_{p^m(i), p^m(j)}}{E_{p^m(i)} E_{p^m(j)}} \right) \tag{25}$$

Since $p^{M_2-1}(i) = p^{M_2-1}(j)$, the first assumption from line (21) applied to line (24) gives

$$\frac{1}{\mathbb{E}(X_{p^{M_2}(i)})} \left(\frac{C_{p^{M_2-1}(i), p^{M_2-1}(j)}}{E_{p^{M_2-1}(i)} E_{p^{M_2-1}(j)}} - \frac{C_{p^{M_2-1}(i), p^{M_2-1}(k)}}{E_{p^{M_2-1}(i)} E_{p^{M_2-1}(k)}} \right) \geq \frac{1}{\mathbb{E}(X_{p^{M_2-1}(i)})} \tag{26}$$

In line (25), note that all summands are non-negative except possibly when $m = M_1 - 1$, therefore

$$\sum_{m=M_1-1}^{M_2-2} \frac{1}{\mathbb{E}(X_{p^{m+1}(i)})}\left(\frac{C_{p^m(i),p^m(j)}}{E_{p^m(i)}E_{p^m(j)}}\right) \geq \frac{1}{\mathbb{E}(X_{p^{M_1}(i)})}\left(\frac{C_{p^{M_1-1}(i),p^{M_1-1}(j)}}{E_{p^{M_1-1}(i)}E_{p^{M_1-1}(j)}}\right) \tag{27}$$

Combining (26) and (27) with (24) and (25), we obtain

$$\begin{align} \frac{\mathrm{Cov}(X_i,X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} - \frac{\mathrm{Cov}(X_i,X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} &\geq \frac{1}{\mathbb{E}(X_{p^{M_2-1}(i)})} + \frac{1}{\mathbb{E}(X_{p^{M_1}(i)})}\left(\frac{C_{p^{M_1-1}(i),p^{M_1-1}(j)}}{E_{p^{M_1-1}(i)}E_{p^{M_1-1}(j)}}\right) \tag{28} \\ &> \frac{1}{\mathbb{E}(X_{p^{M_2-1}(i)})} - \frac{1}{\mathbb{E}(X_{p^{M_1}(i)})} \geq 0 \tag{29} \end{align}$$

where to get line (29) we have applied the second assumption in line (21). This proves "only if" direction. To prove the "if" direction, we must now invoke the assumption from line (22). Suppose that the triplet of leaves $i, j, k \in T$ are all equally related, (meaning $M_1 = M_2$). Then

$$\frac{C_{p^{M_2-1}(i),p^{M_2-1}(j)}}{E_{p^{M_2-1}(i)}E_{p^{M_2-1}(j)}} = \frac{C_{p^{M_2-1}(i),p^{M_2-1}(k)}}{E_{p^{M_2-1}(i)}E_{p^{M_2-1}(k)}} \implies \frac{\mathrm{Cov}(X_i,X_j)}{\mathbb{E}(X_i)\mathbb{E}(X_j)} = \frac{\mathrm{Cov}(X_i,X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} \tag{30}$$

□

Theorem 3 shows that tree reconstruction by neighbor joining is valid when the division and differentiation processes D_i at each node of T follow a set of moment conditions. The next theorem shows that these moment conditions are satisfied for a broad but simple class of division and differentiation processes.

Theorem 4. *Suppose that at each internal node of $i \in T$, cells divide and differentiate by first expanding, and then choosing multinomially between the available daughter cell types. Let α_i be a random variable representing the burst-size distribution for the expansion occurring at node i. Let us assume further that there is at least the possibility for cell division at every node, meaning $P(\alpha_i > 1) > 0$ for all i. Then each resulting D_i distribution satisfies the conditions of Theorem 3 (lines 21 and 22).*

Proof. Let $i, j, k \in T$ be sister nodes, meaning they share a single parent $p(i)$. Let q_i, q_j and q_k denote probability of multinomially choosing each daughter. It follows that

$$\psi_{p(i)}(z_i, z_j, z_k, ...) = \psi_\alpha(p_i z_i + p_j z_j + p_k z_k + ...) \quad \text{where } \psi_\alpha \text{ is the p.g.f. of } \alpha \tag{31}$$

It is now possible to calculate the moments E_i, $C_{i,j}$

$$\begin{align} E_i &= \left.\frac{\partial \psi_{p(i)}}{\partial z_i}\right|_{\mathbf{1}} = \psi'_\alpha(1)p_i = \mathbb{E}(\alpha)p_i \tag{32} \\ C_{i,j} &= \left.\left(\frac{\partial^2 \psi_{p(i)}}{\partial z_i \partial z_j} - \frac{\partial \psi_{p(i)}}{\partial z_i}\frac{\partial \psi_{p(i)}}{\partial z_j} + \delta_{i,j}\frac{\partial \psi_{p(i)}}{\partial z_i}\right)\right|_{\mathbf{1}} \tag{33} \\ &= \psi''_\alpha(1)p_i p_j - (\psi'_\alpha(1))^2 p_i p_j + \delta_{i,j}\psi'_\alpha(1)p_i \tag{34} \\ &= \left(\psi''_\alpha(1)p_i p_j + \psi'_\alpha(1)p_i p_j - (\psi'_\alpha(1))^2 p_i p_j\right) - \psi'_\alpha(1)p_i p_j + \delta_{i,j}\psi'_\alpha(1)p_i \tag{35} \\ &= \mathrm{Var}(\alpha)p_i p_j - \mathbb{E}(\alpha)p_i p_j + \delta_{i,j}\mathbb{E}(\alpha)p_i \tag{36} \end{align}$$

We can now prove the three statements on line (21) and (22). First, it is now trivial that

$$\frac{C_{i,j}}{E_i E_j} = \frac{C_{i,k}}{E_i E_k} \text{ when } j \neq i \neq k \tag{37}$$

For the second statement, observe

$$\frac{C_{i,i}}{E_i^2} - \frac{C_{i,j}}{E_i E_j} = \frac{\delta_{i,i}}{\mathbb{E}(\alpha) p_i} - \frac{\delta_{i,j}}{\mathbb{E}(\alpha) p_i} = \frac{1}{E_i} \implies \frac{C_{i,i}}{E_i^2} \geq \frac{C_{i,j}}{E_i E_j} + \frac{1}{E_i} \tag{38}$$

For the third statement, note that

$$\frac{C_{i,j}}{E_i E_j} \geq \frac{\mathrm{Var}(\alpha)}{\mathbb{E}(\alpha)^2} - \frac{1}{\mathbb{E}(\alpha)} = \frac{\mathbb{E}(\alpha^2) - \mathbb{E}(\alpha)}{\mathbb{E}(\alpha)^2} - 1 > -1 \tag{39}$$

Here we have used the fact that $\mathbb{E}(\alpha^2) > \mathbb{E}(\alpha)$, which follows from the assumption that $\alpha > 1$ with nonzero probability, as well as α being integer valued.

□

5 Detection of tree violations

A simple consequence of Theorem 2 is that the normalized covariances between leaf nodes should be conformally symmetric, in a manner made precise below:

Theorem 5. *Let $i, j, k \in T$ be leaf nodes and suppose that i and j are closer to each other than they are to k, in the sense that there exists M such that $p^M(i) = p^M(j) \neq p^M(k)$. Then*

$$\frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} \tag{40}$$

Proof. Let M' be the minimal integer such as $p^{M'}(i) = p^{M'}(k)$ and M the minimal integer such that $p^M(i) = p^M(j)$. The premise is that $M' - 1 \geq M$. The result follows from theorem 2. □

The property of conformal symmetry can be used as a consistency check for the model. When it is violated, then at least one of the model assumptions must be incorrect. One obvious scenario that would cause a violation of symmetry is if the underlying process cannot be presented as a tree, in the sense that there are multiple paths to same end state. A full accounting of the statistics of multi-type branching on arbitrary directed graphs is outside the scope of this paper. We might reasonably ask, however, how a single tree violation superimposed on an otherwise valid tree would disturb the conformal equalities imposed by Theorem 5. The following theorem specifies which conformal equalities will be violated when a differentiation hierarchy is augmented with a single non-tree transition. To categorize the different types of violations, it will be useful to define distance measure for nodes in the tree. Let the "height" $\mathcal{H}(j)$ refer to the minimum h such that $p^h(i) = j$, where i is a leaf node. And let $\mathcal{S}(j)$ denote the subtree rooted at node j. Then the distance for a pair of nodes i, j is defined

$$d(i,j) = \min\{\mathcal{H}(l) \mid \mathcal{S}(l) \cap \mathcal{S}(i) \neq \emptyset \text{ and } \mathcal{S}(l) \cap \mathcal{S}(j) \neq \emptyset\}$$

Thus, two nodes i and j have distance $d(i,j) = 0$ when one is descended from the other, otherwise their distance is the height of their most recent common ancestor. We can now state how a non-tree transition would affect conformal symmetry.

Theorem 6. *Let T be a tree with parent map p that satisfies the premises of Theorem 4, i.e. in which division and differentiation are independent. Suppose that there exists a single pair of nodes $i', j' \in T$ (not necessarily leaf nodes), where j' is not the ancestor of i', yet cells are allowed to pass directly from j' to i', meaning that $X_{i'}$ is the sum of cells received both from j' and from $p(i')$. Suppose that i, j, k are leaf nodes where conformal symmetry would normally apply (i.e. $d(i,k) = d(j,k) > d(i,j)$). Then a symmetry violation of the form*

$$\frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} > \frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} \tag{41}$$

will be induced under the following conditions

$$\begin{cases} [\text{Case 1};\;\; 0 < d(k,i') < d(k,j')] & \text{violation occurs if and only if} \quad d(i,i') = 0 \text{ and } d(j,i') > 0 \\ [\text{Case 2};\;\; d(k,i') > d(k,j')] & \text{violation occurs if and only if} \quad d(i,i') > 0 \text{ and } d(j,i') = 0 \\ [\text{Case 3};\;\; d(k,i') = 0] & \text{violation occurs if and only if} \quad d(i,i') > d(i,j') > d(j,j') \\ [\text{Case 4};\;\; d(k,i') = d(k,j')] & \text{violation never occurs} \end{cases}$$

Proof. Noting that cases (1-4) are exhaustive, we will consider each in turn. First, let's define new notation. Let X_i^n denote the count of cells in node i that originated from the non-tree transition to i', and let $X_i^t = X_i - X_i^n$ represent the count of cells that arrived by the normal route. When i is not descended from i', $X_i^n = 0$ automatically. It will also be useful to have the following simple facts on hand, which can b easily verified:

Proposition 1:

$$\text{For all } (a,b,c,d,a',b',c',d') \in \mathbb{R}, \;\text{ if }\; \frac{a}{b} = \frac{a'}{b'} \text{ and } \frac{c}{d} = \frac{c'}{d'} \;\text{ then }\; \frac{a+c}{b+d} = \frac{a'+c'}{b'+d'} \tag{42}$$

Proposition 2:

$$\text{For all } (a,b,c,d) \in \mathbb{R}, \;\text{ if }\; \frac{a}{b} < \frac{c}{d} \;\text{ then }\; \frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d} \tag{43}$$

[Case 1] Since the non-tree transition only affects leaf nodes descended from i', violations can only occur when at least one of i, j or k is among the descendants. The premise of case 1 stipulates $d(k,i') > 0$, so k is not. It follows that $d(i,i') = 0$ or $d(j,i') = 0$. Let us assume (without loss of generality) that $d(i,i') = 0$. Since conformal symmetry would normally apply for i, j, k, we have

$$\frac{\mathrm{Cov}(X_j^t, X_k)}{\mathbb{E}(X_j^t)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i^t, X_k)}{\mathbb{E}(X_i^t)\mathbb{E}(X_k)} \tag{44}$$

If $d(j,i') = 0$, then

$$\frac{\mathrm{Cov}(X_j^n, X_k)}{\mathbb{E}(X_j^n)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^n)\mathbb{E}(X_k)} \tag{45}$$

So by proposition 1

$$\frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i^t, X_k) + \mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^t)\mathbb{E}(X_k) + \mathbb{E}(X_i^n)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j^t, X_k) + \mathrm{Cov}(X_j^n, X_k)}{\mathbb{E}(X_j^t)\mathbb{E}(X_k) + \mathbb{E}(X_j^n)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} \tag{46}$$

If, on the other hand, $d(j,i') > 0$, then the premise that $d(k,i') < d(k,j')$ and Theorem 4 imply

$$\frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j^t, X_k)}{\mathbb{E}(X_j^t)\mathbb{E}(X_k)} > \frac{\mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^n)\mathbb{E}(X_k)} \tag{47}$$

So by proposition 2

$$\frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i^t, X_k) + \mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^t)\mathbb{E}(X_k) + \mathbb{E}(X_i^n)\mathbb{E}(X_k)} > \frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} \tag{48}$$

This proves Case 1.

[**Case 2**] Again, at least one of i, j or k must be equal to or descended from i', and the premise excludes k, so either $d(i,i') = 0$ or $d(j,i') = 0$. Let us assume (again without loss of generality) that $d(j,i') = 0$. Since conformal symmetry would normally apply for i, j, k, we have

$$\frac{\mathrm{Cov}(X_i^t, X_k)}{\mathbb{E}(X_i^t)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j^t, X_k)}{\mathbb{E}(X_j^t)\mathbb{E}(X_k)} \tag{49}$$

And by Theorem 4, the additional premise that $d(k,i') > d(k,j')$ implies

$$\frac{\mathrm{Cov}(X_j^n, X_k)}{\mathbb{E}(X_j^n)\mathbb{E}(X_k)} \geq \frac{\mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^n)\mathbb{E}(X_k)} \text{ with equality if and only if } d(i,i') = 0 \tag{50}$$

Hence

$$\frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j^t, X_k) + \mathrm{Cov}(X_j^n, X_k)}{\mathbb{E}(X_j^t)\mathbb{E}(X_k) + \mathbb{E}(X_j^n)\mathbb{E}(X_k)} \geq \frac{\mathrm{Cov}(X_i^t, X_k) + \mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^t)\mathbb{E}(X_k) + \mathbb{E}(X_i^n)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} \tag{51}$$

again with equality if and only if $d(i,i') = 0$. This proves Case 2.

[**Case 3**] If $d(i,i') = d(j,i') = 0$, then Theorem 5 can be applied to the subtree rooted at i', and conformal symmetry is maintained. Therefore we may assume $d(i,i') > 0$, which automatically implies $d(j,i') > 0$ by the assumption that conformal symmetry would normally apply to i, j, k. There are now two scenarios to consider, either $d(i,i') < d(i,j')$ (in which case $d(j,i') < d(j,j')$ automatically), or $d(i,i') > d(i,j')$ (which again would automatically imply $d(j,i') > d(j,j')$). In the first scenario, Theorem 5 implies

$$\frac{\mathrm{Cov}(X_i, X_k^n)}{\mathbb{E}(X_i)\mathbb{E}(X_k^n)} = \frac{\mathrm{Cov}(X_j, X_k^n)}{\mathbb{E}(X_j)\mathbb{E}(X_k^n)} \tag{52}$$

Hence

$$\frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i, X_k^n) + \mathrm{Cov}(X_i, X_k^t)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j, X_k^n) + \mathrm{Cov}(X_j, X_k^t)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} \tag{53}$$

In the second scenario, application of Theorem 4 and Proposition 2 imply that when $d(j,j') < d(i,j')$

$$\frac{\mathrm{Cov}(X_j, X_k)}{\mathbb{E}(X_j)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_j^t, X_k) + \mathrm{Cov}(X_j^n, X_k)}{\mathbb{E}(X_j^t)\mathbb{E}(X_k) + \mathbb{E}(X_j^n)\mathbb{E}(X_k)} > \frac{\mathrm{Cov}(X_i^t, X_k) + \mathrm{Cov}(X_i^n, X_k)}{\mathbb{E}(X_i^t)\mathbb{E}(X_k) + \mathbb{E}(X_i^n)\mathbb{E}(X_k)} = \frac{\mathrm{Cov}(X_i, X_k)}{\mathbb{E}(X_i)\mathbb{E}(X_k)} \tag{54}$$

where the inequality is reversed if $d(j, j') > d(i, j')$, and there is equality if $d(j, j') = d(i, j')$. This proves Case 3.

[Case 4] If $d(i, k) < d(i, i')$, then the assumption that conformal would normally apply to i, j, k means that all three belong to a subtree that excluded i', j' and conformal symmetry is maintained. On the other hand, if $d(i, k) > d(i, i')$ then automatically $d(j, k) > d(j, i')$ which implies that conformal symmetry holds for X_i^t, X_j^t as well as X_i^n, X_j^n, and thus for their sum by Proposition 1. This proves case 4.

□