doi:10.1038/nature20777

Supplementary Methods

Sections:

1. **Barcode distance matrix computation**
2. **Analysis of sources of error in MEMOIR operation, and their impact on reconstruction accuracy**
3. **Simulation of MEMOIR operation for missing cells and sparse trees**
4. **Inference of the switching rates between the two *Esrrb* gene expression states**

1. Barcode distance matrix computation

In this section we describe how we computed the matrix of pairwise barcode distances between cells (Figure 3e). These distances are based on the complete barcoded scratchpad profile of each cell, including both the total number of transcripts, denoted as N, and the fraction of those transcripts that colocalized with scratchpad, denoted as p.

First, to account for statistical fluctuations and errors in the detected fraction of colocalized transcripts, we converted the measured value of p to a distribution of possible values for the colocalization fraction. To account for measurement errors, we imposed a lower bound of 0.04 and an upper bound of 0.96 on the measured value of p. To determine the distribution of possible values of p, we assumed that fluctuations follow a binomial distribution with mean Np and variance $Np(1-p)$. This assumption follows from the physical reasoning that the probability of colocalization of one transcript is independent from that of other transcripts in the cell.

The distribution of possible values of the colocalization fraction, x, takes the form,

$$F(xN) = \binom{N}{x} p^{xN}(1-p)^{(1-x)N}$$

To keep numerical computations tractable, we approximated the factorials in the binomial coefficient using gamma functions, i.e. $N! \approx \Gamma(N+1)$.

To compute the cell-to-cell distance for a single barcode, we quantified how the colocalization fraction distribution differs for a given barcode between two cells. We denote the distribution of colocalization

fraction of barcode i in cell a as $F_a^i(x)$ and the comparable distribution in cell b as $F_b^i(y)$. We define the barcode $d_i(a,b)$ distance between the two cells as,

$$d_i(a,b) = \int_0^1 \int_0^1 F_a^i(x) F_b^i(y) |x-y| dx dy$$

Finally, the overall cell-to-cell distance based on all barcode information was calculated as the average over the distances derived for each barcode weighted for each barcode by the smaller number of barcode transcripts in either of the two cells. This ensures that barcodes that are highly expressed in both cells are weighted more heavily than barcodes that are weakly expressed in either or both cells. Denoting the number of transcripts of barcode i in cell a as N_a^i, and similarly for cell b, the cell-to-cell distance matrix, shown in Fig. 3e, then takes the form,

$$d(a,b) = \sum_i \min\left(N_a^i, N_b^i\right) d_i(a,b)$$

2. Analysis of sources of error in MEMOIR operation, and their impact on reconstruction accuracy

2a. Incorporating noise into MEMOIR simulations

To understand what factors impact the overall reconstruction errors in MEMOIR, we simulated the recording, readout, and reconstruction processes, incorporating relevant sources of stochastic fluctuations (noise). Sources of error can be divided into three categories: (1) noise inherent to the basic mechanism of stochastic scratchpad collapse, (2) recording noise, and (3) readout noise (see Extended Data Fig. 6a).

Noise inherent to the system. The first noise category captures fluctuations due to the general approach of recording lineage information in discrete stochastic mutagenic (collapse) events, but does *not* include noise arising from the specific experimental implementation of the system (those noise sources are described below). Reconstruction errors in this category result from collapse events that by chance occur too frequently or too infrequently to provide lineage information, as well as identical collapse events occurring in different cells (coincidences). For example, if the exact same scratchpad collapse event(s) occur independently in two sister cells, the resulting clades will be indistinguishable. The simulations shown in Figure 3g and Extended Data Figure 8 assume irreversible collapse of an idealized set of binary scratchpads, occurring at a constant rate and therefore incorporate only this

source of error. Analysis of these simulations thus shows how the stochasticity of collapse events impacts reconstruction accuracy in the absence of additional sources of noise.

Recording noise. Recording noise accounts for fluctuations in the collapse rate due to stochastic fluctuations in the expression levels of gRNA and Cas9. To accurately represent the magnitude of these effects, we first measured these fluctuations experimentally, and then incorporated empirical values for these fluctuations into the simulations.

First, we quantified gRNA expression variability both within colonies (intra-colony) and between colonies (inter-colony) using the co-expressed fluorescent reporter in individual cells in all of the 108 MEM-01 colonies. We denote the gRNA expression in cell i in colony m as g_i^m. (Extended Data Fig. 6b,c). The intra-colony noise was computed by first determining the variance of the single-cell measurements amongst cells in each individual colony and then averaging that variance across all colonies. Namely,

$$\sigma_g^2 = \frac{1}{N}\sum_{m=1}^{N}\left[\frac{1}{n_m}\sum_{i=1}^{n_m}\left(g_i^m - \frac{1}{n_m}\sum_{i=1}^{n_m} g_i^m\right)^2\right]$$

Here, N is the total number of colonies, and n_m is the number of cells in colony m.

Similarly, the inter-colony noise was computed by taking the mean of g_i^m amongst the cells in a given colony and then computing the variance of the colony means across all colonies.

$$\tau_g^2 = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{1}{n_m}\sum_{i=1}^{n_m} g_i^m - \bar{g}\right)^2$$

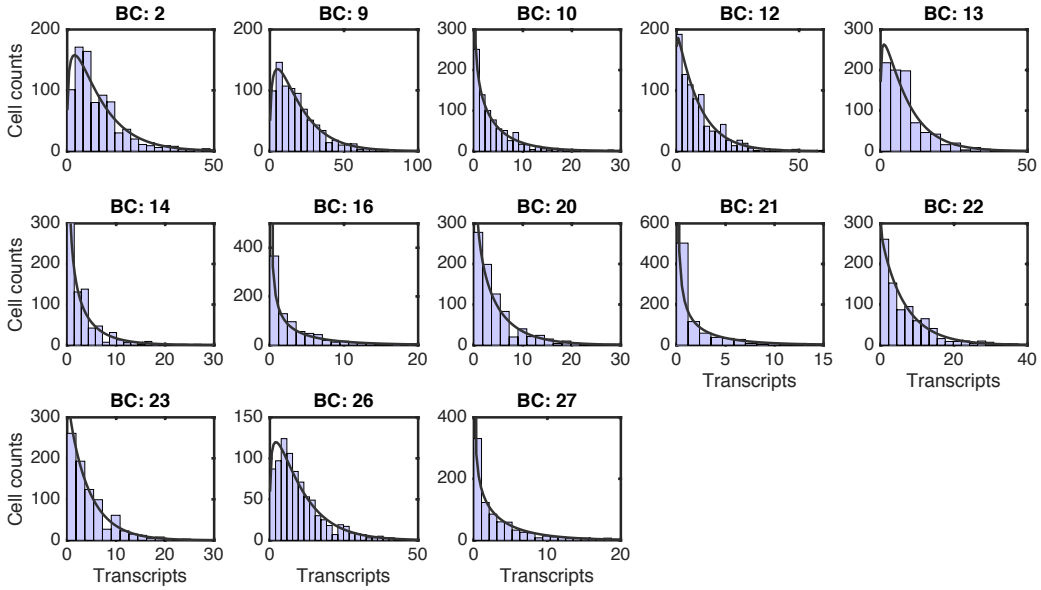
where $\bar{g}$ is the mean gRNA expression across all cells in all colonies,

$$\bar{g} = \frac{1}{N}\sum_{m=1}^{N}\left[\frac{1}{n_m}\sum_{i=1}^{n_m} g_i^m\right]$$

Second, we quantified variability in Cas9 expression using the same analysis, performed on the number of Cas9 transcripts, c, read out using smFISH in the final round of hybridization. We denote the intra-colony and inter-colony variation in Cas9 expression as σ_c and τ_c, respectively, and the mean Cas9 expression level across all cell in all colonies as $\bar{c}$.

We next incorporated the empirically determined noise measurements for Cas9 and gRNA into the simulations. We set the effective collapse rate in each cell, λ, proportional to the product of the number of Cas9 transcripts, c, and the gRNA activity, g: $\lambda = gc$. To incorporate recording noise, we assumed each colony had its own values for c and g, drawn from the observed intercolony distributions with means $\bar{c}$ and $\bar{g}$ and standard deviations τ_c and τ_g respectively. Next, we incorporated the intra-colony noise, by adding fluctuations to values of c and g in each cell in a given colony, with the magnitude of fluctuations set by the empirical intra-colony variations, σ_c and σ_g. The probability of collapse for each scratchpad per generation, p, was calculated from λ assuming a Poisson process, namely, $p = 1 - e^{-\eta\lambda}$. The normalization constant, η, was selected to ensure an average collapse probability of 0.1 per scratchpad per generation, consistent with the empirically estimated average collapse rate in the MEM-01 colonies (see Fig. 2c and Methods). We verified that the resulting fluctuations in the collapse rate (from roughly 0.05 to 0.2 per scratchpad per generation) were consistent with the experimentally observed fluctuations in the loss of colocalization fraction between individual cells across the 108 MEM-01 colonies used in the analysis shown in Figure 3. Overall, as shown in Fig. 3i and Extended Data Fig. 6d, this recording noise produced only a relatively small decrease in the reconstruction accuracy of the simulated colonies.

Readout noise. This category includes (a) stochastic expression of individual barcoded scratchpads, (b) the effects of multiple integrations of the same barcoded scratchpad type, and (c) errors in smFISH imaging readout. To simulate the impact of readout noise on reconstruction accuracy, we first obtained the empirical distribution of single-cell transcript counts of each of the 13 scratchpads in the cells in all 108 analyzed MEM-01 colonies (see Figure below). The transcript count distributions for the barcoded scratchpads were well-approximated as gamma distributions (with an average mean of 5.4). The gamma distribution is expected when there is a constant probability per unit time of initiating a transcriptional burst, and the number of mRNAs produced per burst follows an exponential distribution, and has been observed empirically to describe diverse stochastic gene expression processes[43,44].



Distribution of the number of transcripts of each barcode in individual cells in the 108 MEM-01 colonies analyzed in Figure 3. All 13 detected barcodes are shown. Black lines are fits to Gamma distributions.

We quantified the intra-colony and inter-colony variability in the expression levels of the barcoded scratchpads using the same procedure described above for gRNA and Cas9. In this case, variance was calculated for each barcode and then summed over 13 barcodes per cell to estimate each cell's overall expression variance. These empirical parameterizations for the transcript count distributions were incorporated in the simulations.

For readout, we first associated with each barcode a random number of transcripts drawn from the gamma distribution associated with that barcode. Second, we incorporated additional noise from smFISH imaging, based on experimental quantification of the accuracy of smFISH colocalization detection rates (Fig. 2c). If a scratchpad was intact, then the probability of colocalization, p_c, was set to 0.87; if a scratchpad was collapsed, the probability of erroneous colocalization, p_e, was set to 0.05. These estimates were obtained from our measurements of colocalization fractions before activation of the MEM-01 cells (Fig. 2c), and are consistent with previous reports on smFISH fidelity[16,17]. The number of scratchpads that colocalized for each barcode was drawn from binomial distributions based on these colocalization detection probabilities. Namely, the number of colocalized transcripts q from a total of T transcripts, for an intact integration of a barcode, was drawn from the distribution,

$$p_{intact}(q;T) = \binom{T}{q} p_c^q (1-p_c)^{T-q}$$

Similarly, the number of colocalized transcripts for a collapsed integration of a barcode was drawn from the distribution,

$$p_{collapsed}(q;t) = \binom{T}{q} p_e^q (1-p_e)^{T-q}$$

Finally, we accounted for multiple integrations of the same barcode as follows: If a barcode had more than one integration, the total number of transcripts and the number of colocalized transcripts for that barcode were determined by summing the relevant transcripts from each individual integration. Importantly, the distribution of the expression level of each individual integration was selected such that the distribution of the total number of transcripts (obtained by taking the convolution of the single-integration distributions) matched the experimentally observed distribution for that barcode. For example, if a barcode had two integrations and an expression distribution characterized by a mean of 10 transcripts and shape parameter 2 (from the gamma fit), then each integration was assumed to have an expression distribution characterized by a mean of 5 transcripts and shape parameter 1.

This procedure produced an ensemble of simulated colonies, where every cell was assigned a total number of transcripts and a number of colocalized transcripts for each barcode (same as the experimental data). The resulting simulated colonies were then reconstructed with the same neighbor-joining algorithm used on the experimental data (see Methods).

Impact of readout noise on overall reconstruction accuracy. Figure 3i and Extended Data Figure 6e,f show that noise in scratchpad expression and multiple incorporations of the same barcoded scratchpad account for most of the reconstruction error. By contrast, other components such as recording noise, noise intrinsic to the MEMOIR design, and noise from smFISH imaging fidelity contributed minimally to the overall reconstruction error.

Finally, to determine if the addition of fluctuations to the simulations matched the experimentally observed decrease in reconstruction accuracy, we simulated trees with all three components of noise for various numbers of integrations per barcode and compared the distribution of reconstruction accuracy to that observed for the 108 experimental colonies. As shown in Extended Data Figure 6f, the distribution of reconstruction accuracy obtained from simulations is generally consistent with the experimentally observed distribution, with an effective value of 2 integration sites for each barcode. Note that no fitting parameters were used besides the empirically measured fluctuations in the recording and readout components. Together, these results indicate that this implementation of MEMOIR behaved as expected of a system with 13 scratchpads and realistic sources of noise. The main sources of error, the stochastic expression of scratchpad transcripts and variable expression from multiple integrations of the same barcoded scratchpad types, can be improved in future implementations.

2b. Potential solutions to address sources of noise and reduce errors

Here we describe ways to mitigate the key sources of noise currently affecting MEMOIR reconstruction accuracy:

- **Noise inherent to the system.** Because collapse events are stochastic, similar collapse patterns can arise by chance independently, creating ambiguities. The likelihood of such errors decreases rapidly as the number and diversity of barcoded scratchpads is increased. Further, tunable control of Cas9 and gRNA expression or scratchpad affinity (e.g., by varying gRNA target complementarity) allows average collapse rate to be optimized for different regimes, including accessing more generations (Extended Data Fig. 8b) or low collapse rates for studying sparse trees (Extended Data Fig. 9).
- **Readout noise: Multiple incorporations of barcoded scratchpad types.** This is currently a significant source of noise, as shown in Extended Data Fig. 6, and it is readily addressed as above, e.g., by starting from a larger and more diverse pool of barcodes.
- **Readout noise: Stochastic scratchpad expression.** Individual scratchpads are read out at the RNA level, and therefore subject to fluctuations in transcription. This source of noise that can be addressed by using 1) stronger promoters, 2) site-specific integration in well-expressed loci, and/or 3) stabilized RNA transcripts (in order to build up signal over longer periods). Alternatively, barcode expression noise can be circumvented by employing other in situ labeling methods, like single-molecule DNA FISH.

3. Simulation of MEMOIR operation for missing cells and sparse trees

3a. Missing cells

To simulate reconstruction for trees with missing cells (Extended Data Fig. 8c), the forward algorithm described in Methods was used to generate a full binary tree of three generations, which was then pruned by removing a given number of randomly chosen endpoint cells. To reconstruct the lineage relationships of the remaining cells, we assigned each of the possible 315 reconstructions a score given by $\sum_{<i,j>}\left(H_{ij}-T_{ij}\right)^2$, where H_{ij} is the Hamming distance between the barcoded scratchpad collapse patterns of cells i and j, T_{ij} is the lineage distance between cells i and j from that particular reconstruction (1 for sisters, 2 for first cousins, and 3 for second cousins), and the summation runs over all pairs of cells. Matrices H and T were normalized so that their elements summed to one. This reconstruction score format was useful not only in scoring the topology of possible lineage arrangements, but also in estimating branch lengths (e.g., whether two cells were likely closely related sisters or more distantly related cousins). The reconstruction with the lowest score was selected as the reconstructed tree and used to compute the fraction of relationships that were correctly identified (i.e., fraction of correct relationships) shown in the heat-maps. If multiple reconstructions had the same lowest score, the fraction of correct relationships was averaged over all of them.

3b. Simulation of MEMOIR operation for sparse trees

To simulate sparse trees, we assumed collapse events in a given lineage occur at a specified constant collapse rate λ (per cell per generation). The sparse regime occurs when λ < 1. Because generating hundreds of collapse events requires large trees of many generations, we did not simulate the underlying proliferating binary tree. Instead, we directly generated a tree of collapse events with variable branch lengths corresponding to the stochastic time intervals between collapse events. Assuming a constant rate of collapse (a Poisson process), we randomly drew branch lengths from an exponential distribution with the time scale set by the collapse rate, i.e. mean branch length ~ 1/λ . At any given level of the tree, if the tree had N branches, the time interval to the next branching event was selected from an exponential distribution with rate $N\lambda$, and the branching event was randomly assigned to one of the N branches. The branching process was continued until the desired number of leaves was generated. Extended Data Figure 9c,d shows examples of sparse trees with various number of leaves and tree depths (defined as the cumulative number of collapse events since the common ancestor experienced by each leaf averaged over all the leaves of the tree). For clarity, the approximate

correspondence between these sparse trees (characterized by number of leaves and tree depth) and underlying full binary trees (characterized by collapse rate and number of generations) was determined by simulating full binary trees in the sparse regime and then finding parameters that generated effective sparse trees with the desired number of leaves and depth (see the cartoon in Extended Data Fig. 9a). The collapse rate and number of generations of the full binary trees corresponding to the example sparse trees are reported in Extended Data Figure 9c,d. To reconstruct sparse trees, we used the generated collapse pattern at the leaves and a Sankoff maximum parsimony algorithm[45]. The fraction of correct partitions identified by the reconstructed trees was calculated using the Robinson-Foulds distance metric[40]. Simulations were implemented in Matlab, and Python using the Biopython package[46] and ETE3 toolkit[47].

4. Inference of the switching rates between the two *Esrrb* gene expression states

In this section, we describe how we inferred the transition rates between the low and high expression states of *Esrrb*. In 4a, we compute the occurrence frequencies of pairs of cells in the same gene expression state. In 4b, we present a straightforward inference of the transition rates from the decay in the observed frequency of pairs of cells in the same expression state. Finally, in 4c, we present a more involved but more accurate inference algorithm.

4a. Occurrence frequency of related pairs of cells in the same *Esrrb* gene expression state on MEMOIR trees

In this subsection, we describe how the co-occurrence frequencies of pairs of related cells in the same *Esrrb* expression state can be extracted from the MEMOIR reconstructed lineage trees and endpoint gene expression measurements.

To compute the frequencies, we first assign each cell a probability of being in one of the two *Esrrb* gene expression states based on its *Esrrb* mRNA count. To do so, we fit the bimodal distribution of single-cell *Esrrb* transcript counts (Figure 4b) with a weighted sum of two negative binomial distributions, each corresponding to one gene expression state. We denote the distribution corresponding to the *Esrrb* low expression state as $P_{-}(m)$, and the distribution corresponding to the *Esrrb* high expression state as $P_{+}(m)$, where m is the number of transcripts in an individual cell. The bimodal distribution is fit to $fP_{-}(m) + (1-f)P_{+}(m)$, where f is an additional fitting parameter that corresponds to the population fraction of cells in the low expression state.

With the fit parameters determined, the probability that an individual cell with n *Esrrb* mRNA molecules is in the low expression state (E-) is

$$p_{E-}(n) = \frac{fP_{-}(n)}{fP_{-}(n) + (1-f)P_{+}(n)}$$

Similarly, the probability that the cell is in the high gene expression state (E+) is simply $p_{E+}(n) = 1 - p_{E-}(n)$.

Next, we generalize this approach to jointly observing the states of a pair of cells. If cell 1 is in the E- state with probability p^1_{E-} and cell 2 is in the E- state with probability p^2_{E-} then the probability that both cells are simultaneously in E- states is $p^1_{E-}p^2_{E-}$. Similarly, the probability of observing cell 1 in the E+ state and cell 2 in the E- state is $p^1_{E+}p^2_{E-}$. We can denote the probability of observing the two cells in any of the four possible pairs of states by a 2×2 matrix,

$$\begin{bmatrix} p^1_{E-}p^2_{E-} & p^1_{E+}p^2_{E-} \\ p^1_{E-}p^2_{E+} & p^1_{E+}p^2_{E+} \end{bmatrix}$$

The entries in this matrix must sum to 1. Finally, to compute the occurrence frequency of a pair of sisters cells in each of the four possible pairs of states, we average the above matrix over all pairs of sisters cells.

$$\boldsymbol{C}(1) = \frac{1}{N_{sisters}} \sum_{\langle i,j \rangle} \begin{bmatrix} p^i_{E-}p^j_{E-} & p^i_{E+}p^j_{E-} \\ p^i_{E-}p^j_{E+} & p^i_{E+}p^j_{E+} \end{bmatrix}$$

where the summation runs over all $N_{sisters}$ pairs of sisters cells in the 30 colonies used in the analysis. We note that since the ordering of the sister pair is arbitrary, we turn the resulting matrix symmetric by averaging the off-diagonal terms.

$\boldsymbol{C}(1)$ is a 2x2 matrix for sisters. The argument '1' denotes the fact that sisters are 1 generation apart. Similarly, we define $\boldsymbol{C}(2)$ as the frequency of observing a pair of first-cousin cells in a given pair of

states. $\boldsymbol{C}(2)$, is determined by taking the average over all the observed pairs of first-cousin cells. The occurrence frequency of sisters, first cousins, and second cousins ($\boldsymbol{C}(3)$) in the same *Esrrb* expression state (either both E-, or both E+) is plotted in Figure 4d.

Finally, we estimated the statistical error on the observed frequencies in two ways (error bars in Figure 4d). First, from the total number of observed pairs of cells and the computed frequencies, we estimated the counting error assuming a multinomial distribution. That is, for an observed frequency f computed using N pairs of cells, the statistical error was estimated as $\sqrt{f(1-f)N}$. To understand how much of the similarity observed between states of closely related cells was just due to chance, we randomly permuted the lineage relationships within each colony and recomputed the frequencies. The statistical error was estimated as the variance of the frequencies computed over 1000 iterations of such random permutations. The first error estimate only accounts for the finite size of the total number of observations. The second estimate accounts for correlations between the states of closely related cells due to chance with the number of different gene expression states in each colony held fixed. The two methods produced similar estimates reflecting the level intra-colony heterogeneity. Because the two approaches do not capture independent statistical fluctuations, we combined them by taking the maximum of the two estimates.

4b. Inference of switching dynamics from reconstructed lineage trees and endpoint gene expression measurements

Here, we describe the inference of transition rates from the frequencies derived in the previous section. To do so, we first derive the relationship between transitions rates and the expected probability of observing a pair of related cells in the same gene expression state.

We define a minimal model of stochastic and memoryless transitions between the two expression states. We assume that the probability of transitioning from the E- state to the E+ state per generation is constant and denote it as $T(E^+|E^-)$. The probability that a cell in the E- state will retain its state from one generation to the next is give by $T(E^-|E^-) = 1 - T(E^+|E^-)$. Similarly, the probability of transitioning out of the E+ state per generation is denoted as $T(E^-|E^+)$. The transition probabilities can be combined together into a transition rate matrix,

$$\boldsymbol{T} = \begin{bmatrix} T(E^-|E^-) & T(E^-|E^+) \\ T(E^+|E^-) & T(E^+|E^+) \end{bmatrix}$$

First, we present a simple approach for inferring the transition rate matrix from the decay in the observed frequencies of pairs of cells in the same gene expression state going from first-cousins to second-cousins. Next, we present a more involved derivation that also takes into account the value of the observed frequencies in addition to how they decay from closely related cells to more distant relatives.

We can relate the frequency of observing second-cousins in the same state (either both in E- or both E+) to the frequency of observing the first-cousins in a particular pair of states by taking into account all possible state transitions going from one generation to the next. It follows,

$$\boldsymbol{C}_{11}(3) = (1-T_{12})(1-T_{12})\boldsymbol{C}_{11}(2) + 2T_{21}(1-T_{12})\boldsymbol{C}_{12}(2) + T_{21}T_{21}\boldsymbol{C}_{22}(2)$$
$$\boldsymbol{C}_{22}(3) = (1-T_{21})(1-T_{21})\boldsymbol{C}_{22}(2) + 2T_{12}(1-T_{21})\boldsymbol{C}_{21}(2) + T_{12}T_{12}\boldsymbol{C}_{11}(2)$$

We numerically solved the above two non-linear equations for T_{12} and T_{21} for the observed values of $\boldsymbol{C}(2)$ and $\boldsymbol{C}(3)$ (the observed frequencies were first corrected for misclassifications of the two states caused by measurement uncertainties, see below). To estimate the statistical error on the inferred transitions rates, we produced simulated frequency matrices based on the estimated statistical error of the frequencies (see previous section). We solved for the transition rates for 10000 iterations of the simulated data and computed the standard deviation over the rates. We found $T_{12} = 0.10 \pm 0.07$ and $T_{21} = 0.04 \pm 0.03$.

4c. A more accurate inference of the transition rates

Next, to get a more accurate set of inferred transition rates, we incorporated the values of the frequencies in addition to their decay over the generations in our calculations. To do so, we first calculated the joint-probability of finding a pair of sister cells in states i and j using the transition rate matrix.

$$\boldsymbol{C}_{ij}(1) = \sum_{s=E^-,E^+} \boldsymbol{T}(i|s)\boldsymbol{T}(j|s)p(s)$$

The summation in above equation runs over the two possible states of the parent cell. Assuming that the population fractions of the E- and E+ cells do not change significantly during the three generations spanned by the experiment, $p(E^-)$ and $p(E^+)$ can be set approximately equal to the population fractions derived from fitting the bimodal *Esrrb* transcript-count distribution (see previous section). Namely, $p(E^-) = f$ and $p(E^+) = 1 - f$. We experimentally validated the assumption that the *Esrrb* transcript-count distribution does not change significantly in our experimental conditions over a period of 48 hours or approximately four generations (see Extended Data Fig. 10).

To keep the population fractions constant, the number of cells transitioning from the E- state to the E+ state must be equal to the number transitioning in the reverse direction. Hence, $T(E^+|E^-)p(E^-) = T(E^-|E^+)p(E^+)$. With this simplification,

$$\boldsymbol{C}_{ij}(1) = \sum_{s=E^-,E^+} \boldsymbol{T}(i|s)\boldsymbol{T}(s|j)p(j) = p(j)\boldsymbol{T}^2(i|j)$$

where $\boldsymbol{T}^2(i|j)$ denotes the i,jth element of the square of the transition matrix. Similarly, the joint-probability of observing a pair of cousins, and second-cousins in states i and j is respectively equal to $\boldsymbol{C}_{ij}(2) = p(j)\boldsymbol{T}^4(i|j)$ and $\boldsymbol{C}_{ij}(3) = p(j)\boldsymbol{T}^6(i|j)$.

To infer the state transition rates, we used the measured frequencies and population fractions and solved the above equations for $\boldsymbol{T}$.

However, we first need to introduce a correction to the measured frequencies to account for misclassification of expression states when they are probabilistically assigned. Namely, it is more likely to misclassify a cell in the E- state as E+ than vice versa because of the asymmetrical shape of the distributions $P_{\pm}(m)$ (see the fits in Figure 4b). To correct for misclassifications, we first computed the probability of assigning a cell in state j mistakenly to state i by simply integrating the probability that a a cell with n *Esrrb* transcripts is assigned to state i (computed in the previous section) over the probability distribution of observing n transcripts in a cell in state j.

$$\boldsymbol{Q}_{ij} = \int \frac{fP_i(n)}{fP_-(n) + (1-f)P_+(n)} P_j(n)dn$$

For our analysis, E- cells are mistakenly assigned to E+ cells with probability 0.12, whereas E+ cells are mistakenly assigned to the E- state with probability 0.04.

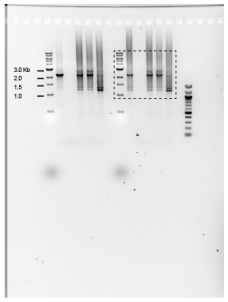
Matrix $\boldsymbol{Q}$ effectively introduces apparent transitions between the two states at the measurement stage. These additional transitions can be explicitly included in the calculation of the joint-probability of finding a pair of sister cells in states i and j,

$$\boldsymbol{C}_{ij}(1) = \sum_{s=E^-,E^+} \sum_{\boldsymbol{k}} \sum_{\boldsymbol{l}} \boldsymbol{Q_{ik}} \boldsymbol{T}(k|s) \boldsymbol{Q_{jl}} \boldsymbol{T}(l|s) p(s) = \sum_{s=E^-,E^+} \sum_{\boldsymbol{k}} \sum_{\boldsymbol{l}} \boldsymbol{Q_{ik}} \boldsymbol{T}(k|s) \boldsymbol{T}(s|l) \boldsymbol{Q_{lj}^T} p(l)$$
$$= \sum_{\boldsymbol{k}} \sum_{\boldsymbol{l}} \boldsymbol{Q_{ik}} \boldsymbol{T^2}(k|l) p(l) \boldsymbol{Q_{lj}^T}$$

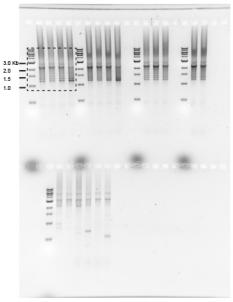
However, the apparent transitions represented by $\boldsymbol{Q}$ do not correspond to actual transitions between the *Esrrb* expression states. Rather, they represent misclassifications due to uncertainties in the state assignment. To correct the estimate, we removed the contributions of these apparent transitions to the measured frequencies, by applying the inverse of matrix $\boldsymbol{Q}$ to the frequency matrices, $\widetilde{\boldsymbol{C}} = \boldsymbol{Q}^{-1}\boldsymbol{C}(\boldsymbol{Q}^T)^{-1}$.

Finally, to infer the transition rates shown in Figure 4d, we solved for the transition rate matrix $\boldsymbol{T}$ using the corrected frequency matrix for second-cousins , $\widetilde{\boldsymbol{C}}(3)$. $\widetilde{\boldsymbol{C}}(3)$ exhibits a lower statistical error than $\widetilde{\boldsymbol{C}}(2)$ and $\widetilde{\boldsymbol{C}}(1)$ because the number of pairs of second cousins is larger than the number of pairs of first-cousins and sisters. To validate the inferred rates, we checked that inferring $\boldsymbol{T}$ using the corrected frequency matrices for first-cousins $\widetilde{\boldsymbol{C}}(2)$ and sisters $\widetilde{\boldsymbol{C}}(1)$ produced consistent transition rates. Finally, to estimate the statistical error of the inferred rates we used the estimated statistical error in the measured frequencies (see previous section) and produced simulated frequency matrices. The error on the inferred rates was estimated as the standard deviation of the rates inferred over 10000 iterations of simulated data. Results are shown in Figure 4d.

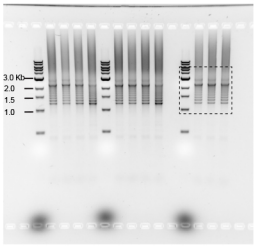
Extended Data Figure 2a



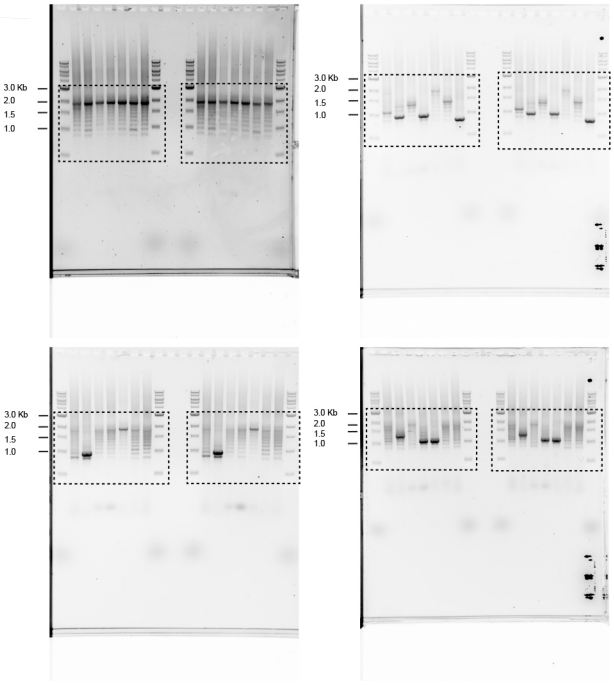
Extended Data Figure 2c



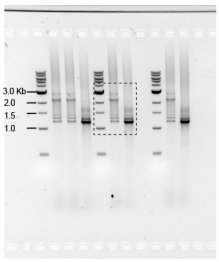
Extended Data Figure 2e



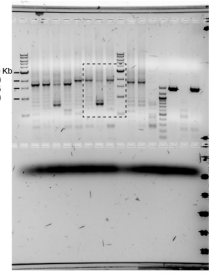
Extended Data Figure 2f



Extended Data Figure 3b



Extended Data Figure 3e



Extended Data Figure 3f