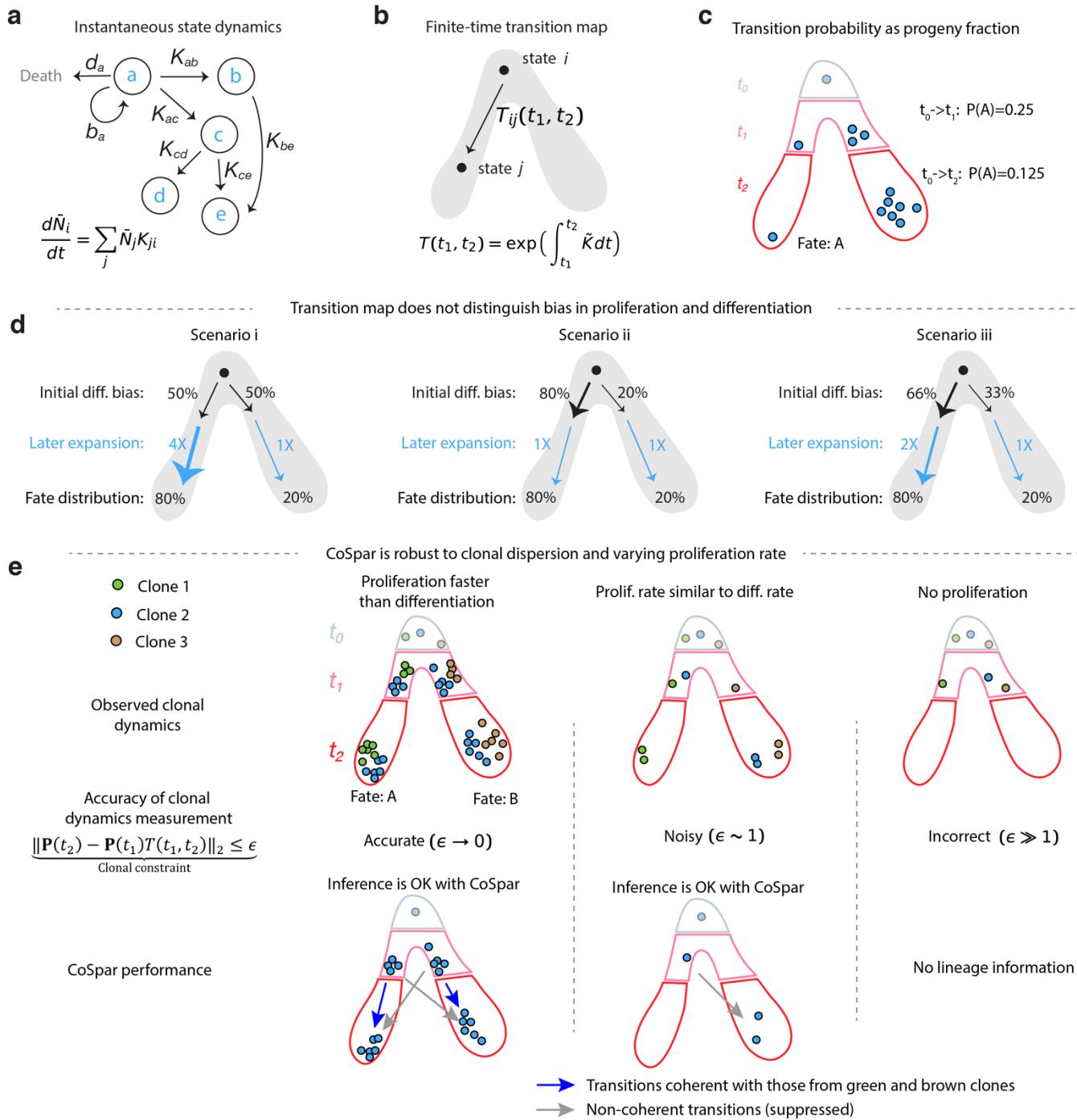


1  
2      *Supplementary Information for*

**Learning dynamics by computational integration of single cell genomic and lineage  
information**

Shou-Wen Wang\* and Alon M. Klein\*

\*Corresponding authors: shouwen\_wang@hms.harvard.edu (S.W.W.), alon\_klein@hms.harvard.edu (A.M.K.)



1

## Supplementary Fig. 1. Models, assumptions and limitations of Coherent Sparse Optimization.

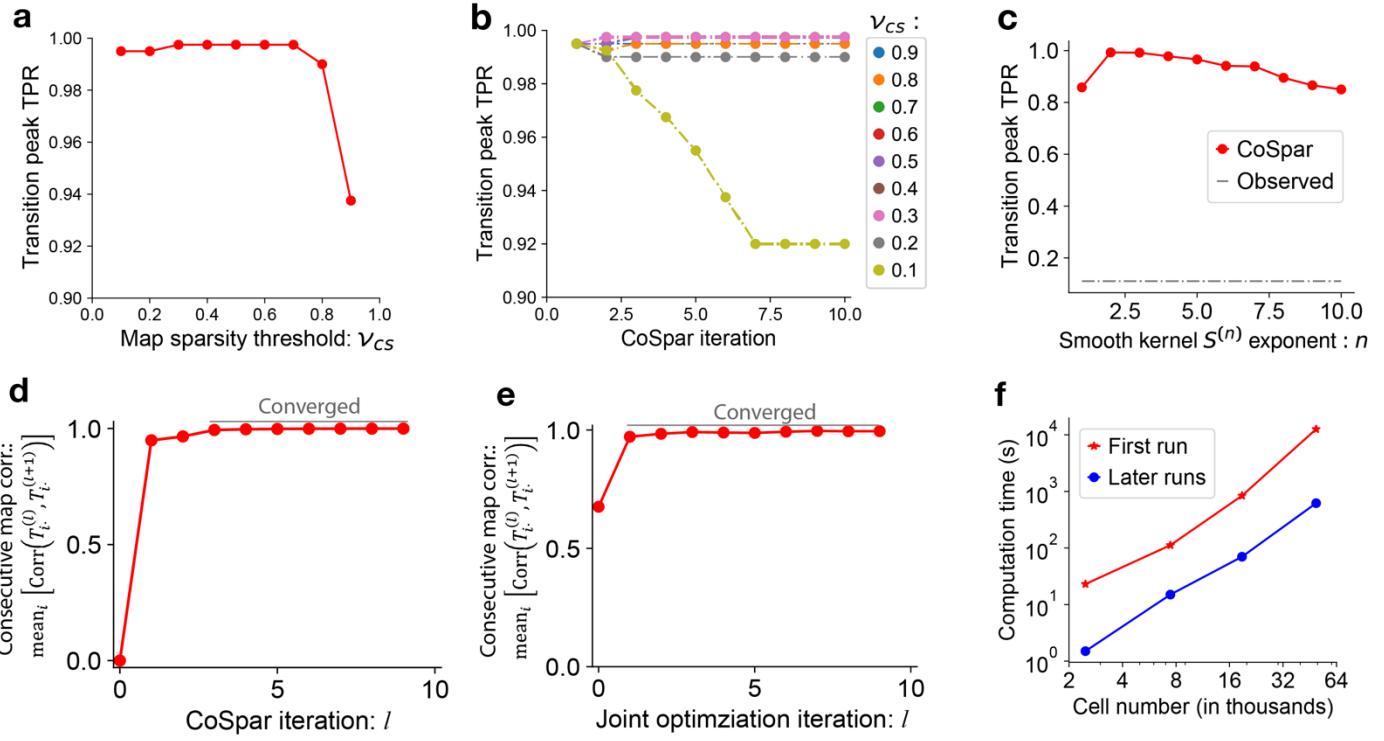
2 **a**, Simple example of the class of stochastic models that CoSpar seeks to learn. In such models, 3 each node represents an observed cell state. In practice, thousands of measured states are included; 4 here only five are shown. At each state cells self-renew, die, or differentiate with state-specific rates. 5 The mean fraction of cells in each state evolves according to coupled first-order equations as shown. 6 See Supplementary Note 1 for details.

7 **b**, The empirically-observed finite-time transition map can be interpreted through its relation to the 8 transition rate matrix  $K$  (see panel **a**). See Supplementary Note 1 for details.

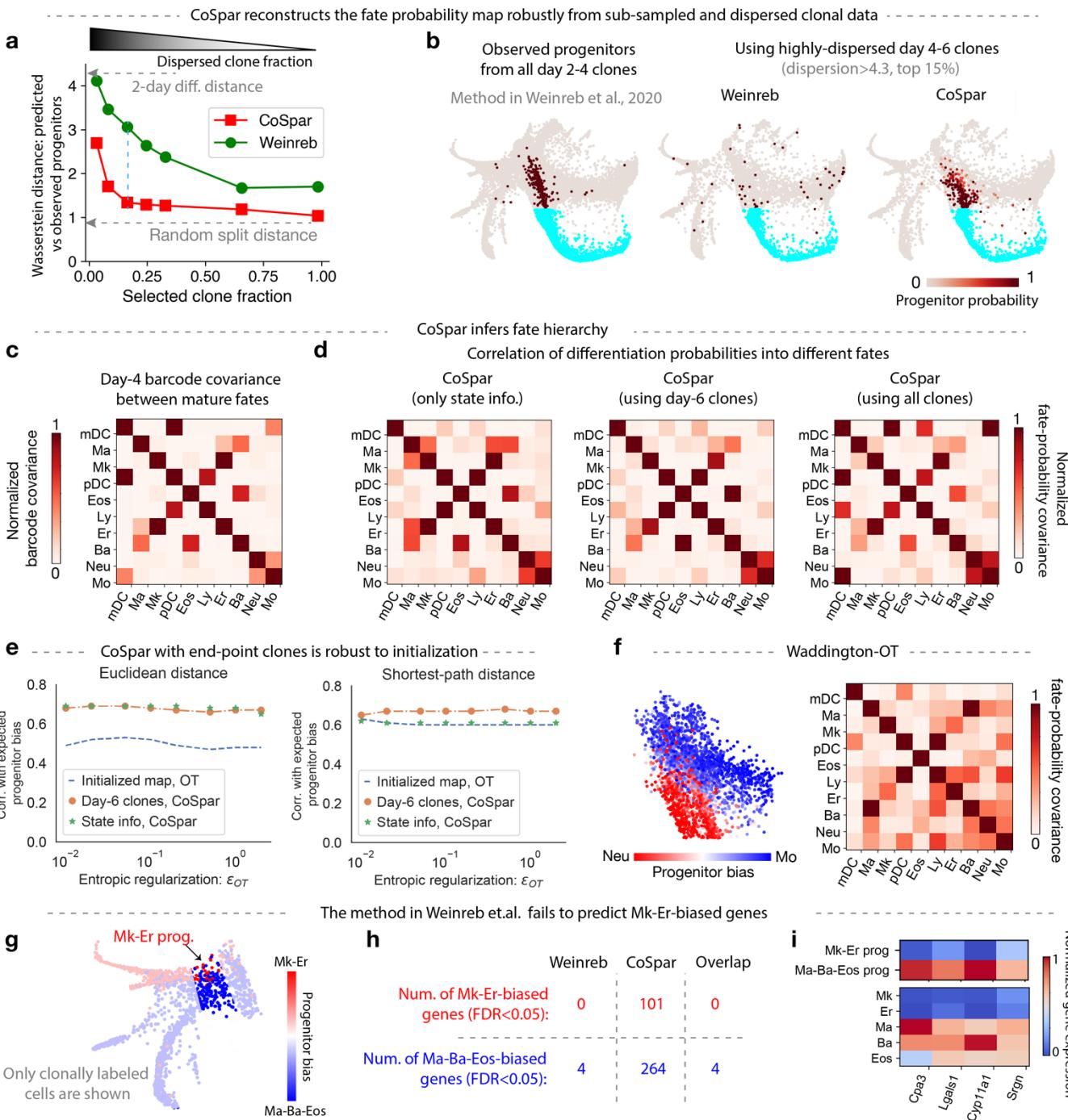
0

1 **c**, Schematics illustrating the operational, experimentally-accessible definition of a transition  
2 probability, as the average fraction of progeny derived from an initial cell  $i$  at  $t_0$  that differentiates into  
3 a target state  $j$  at later times. As defined, transition probabilities are sensitive to biases in fate choice,  
4 and to differential rates of cell division and cell loss.  
5 **d**, Schematics exemplifying that transition maps cannot distinguish fate bias from differences in net  
6 rates of cell expansion (division – loss). Three different underlying dynamics lead to the same  
7 transition maps.  
8 **e**, Schematics clarifying the robustness of CoSpar to clonal dispersion (demonstrated in Fig. 3). i),  
9 When cells undergo extensive proliferation prior to fate bifurcation and clonal sampling, each clone  
0 densely samples several differentiation trajectories. By imposing sparsity and coherence, CoSpar re-  
1 enforces a minimal number of transitions that explain dynamics across all clones. ii), At lower rates of  
2 proliferation, fewer cells from each clone are sampled, and it may lead to observing clonally-related  
3 cells at different time-points on different trajectories, as shown (blue clone sampled towards fate A at  
4  $t_1$ , and towards fate B at  $t_2$ ). By enforcing coherence between clones rooted in neighboring states,  
5 CoSpar may still recover a correct transition map. In this case, there is a trade-off in the CoSpar cost  
6 function between minimizing the clone transition map error and maximizing coherence. iii), Lacking  
7 proliferation, one cannot establish clonal relationships that constrain dynamic inference.

8  
9



## 1      2      3      4      5      6      7      8      9      10      11      12      13      14      15      16      17      18      19      20      21      22      23      24      25      26      27      28      29      30      31      32      33      34      35      36      37      38      39      40      41      42      43      44      45      46      47      48      49      50      51      52      53      54      55      56      57      58      59      60      61      62      63      64      65      66      67      68      69      70      71      72      73      74      75      76      77      78      79      80      81      82      83      84      85      86      87      88      89      90      91      92      93      94      95      96      97      98      99      100      101      102      103      104      105      106      107      108      109      110      111      112      113      114      115      116      117      118      119      120      121      122      123      124      125      126      127      128      129      130      131      132      133      134      135      136      137      138      139      140      141      142      143      144      145      146      147      148      149      150      151      152      153      154      155      156      157      158      159      160      161      162      163      164      165      166      167      168      169      170      171      172      173      174      175      176      177      178      179      180      181      182      183      184      185      186      187      188      189      190      191      192      193      194      195      196      197      198      199      200      201      202      203      204      205      206      207      208      209      210      211      212      213      214      215      216      217      218      219      220      221      222      223      224      225      226      227      228      229      230      231      232      233      234      235      236      237      238      239      240      241      242      243      244      245      246      247      248      249      250      251      252      253      254      255      256      257      258      259      260      261      262      263      264      265      266      267      268      269      270      271      272      273      274      275      276      277      278      279      280      281      282      283      284      285      286      287      288      289      290      291      292      293      294      295      296      297      298      299      300      301      302      303      304      305      306      307      308      309      310      311      312      313      314      315      316      317      318      319      320      321      322      323      324      325      326      327      328      329      330      331      332      333      334      335      336      337      338      339      340      341      342      343      344      345      346      347      348      349      350      351      352      353      354      355      356      357      358      359      360      361      362      363      364      365      366      367      368      369      370      371      372      373      374      375      376      377      378      379      380      381      382      383      384      385      386      387      388      389      390      391      392      393      394      395      396      397      398      399      400      401      402      403      404      405      406      407      408      409      410      411      412      413      414      415      416      417      418      419      420      421      422      423      424      425      426      427      428      429      430      431      432      433      434      435      436      437      438      439      440      441      442      443      444      445      446      447      448      449      450      451      452      453      454      455      456      457      458      459      460      461      462      463      464      465      466      467      468      469      470      471      472      473      474      475      476      477      478      479      480      481      482      483      484      485      486      487      488      489      490      491      492      493      494      495      496      497      498      499      500      501      502      503      504      505      506      507      508      509      510      511      512      513      514      515      516      517      518      519      520      521      522      523      524      525      526      527      528      529      530      531      532      533      534      535      536      537      538      539      540      541      542      543      544      545      546      547      548      549      550      551      552      553      554      555      556      557      558      559      560      561      562      563      564      565      566      567      568      569      570      571      572      573      574      575      576      577      578      579      580      581      582      583      584      585      586      587      588      589      590      591      592      593      594      595      596      597      598      599      600      601      602      603      604      605      606      607      608      609      610      611      612      613      614      615      616      617      618      619      620      621      622      623      624      625      626      627      628      629      630      631      632      633      634      635      636      637      638      639      640      641      642      643      644      645      646      647      648      649      650      651      652      653      654      655      656      657      658      659      660      661      662      663      664      665      666      667      668      669      670      671      672      673      674      675      676      677      678      679      680      681      682      683      684      685      686      687      688      689      690      691      692      693      694      695      696      697      698      699      700      701      702      703      704      705      706      707      708      709      710      711      712      713      714      715      716      717      718      719      720      721      722      723      724      725      726      727      728      729      730      731      732      733      734      735      736      737      738      739      740      741      742      743      744      745      746      747      748      749      750      751      752      753      754      755      756      757      758      759      7510      7511      7512      7513      7514      7515      7516      7517      7518      7519      7520      7521      7522      7523      7524      7525      7526      7527      7528      7529      7530      7531      7532      7533      7534      7535      7536      7537      7538      7539      7540      7541      7542      7543      7544      7545      7546      7547      7548      7549      7550      7551      7552      7553      7554      7555      7556      7557      7558      7559      75510      75511      75512      75513      75514      75515      75516      75517      75518      75519      75520      75521      75522      75523      75524      75525      75526      75527      75528      75529      75530      75531      75532      75533      75534      75535      75536      75537      75538      75539      75540      75541      75542      75543      75544      75545      75546      75547      75548      75549      75550      75551      75552      75553      75554      75555      75556      75557      75558      75559      75560      75561      75562      75563      75564      75565      75566      75567      75568      75569      75570      75571      75572      75573      75574      75575      75576      75577      75578      75579      75580      75581      75582      75583      75584      75585      75586      75587      75588      75589      75590      75591      75592      75593      75594      75595      75596      75597      75598      75599      755100      755101      755102      755103      755104      755105      755106      755107      755108      755109      755110      755111      755112      755113      755114      755115      755116      755117      755118      755119      755120      755121      755122      755123      755124      755125      755126      755127      755128      755129      755130      755131      755132      755133      755134      755135      755136      755137      755138      755139      755140      755141      755142      755143      755144      755145      755146      755147      755148      755149      755150      755151      755152      755153      755154      755155      755156      755157      755158      755159      755160      755161      755162      755163      755164      755165      755166      755167      755168      755169      755170      755171      755172      755173      755174      755175      755176      755177      755178      755179      755180      755181      755182      755183      755184      755185      755186      755187      755188      755189      755190      755191      755192      755193      755194      755195      755196      755197      755198      755199      755200      755201      755202      755203      755204      755205      755206      755207      755208      755209      755210      755211      755212      755213      755214      755215      755216      755217      755218      755219      755220      755221      755222      755223      755224      755225      755226      755227      755228      755229      755230      755231      755232      755233      755234      755235      755236      755237      755238      755239      755240      755241      755242      755243      755244      755245      755246      755247      755248      755249      755250      755251      755252      755253      755254      755255      755256      755257      755258      755259      755260      755261      755262      755263      755264      755265      755266      755267      755268      755269      755270      755271      755272      755273      755274      755275      755276      755277      755278      755279      755280      755281      755282      755283      755284      755285      755286      755287      755288      755289      755290      755291      755292      755293      755294      755295      755296      755297      755298      755299      755300      755301      755302      755303      755304      755305      755306      755307      755308      755309      755310      755311      755312      755313      755314      755315      755316      755317      755318      755319      755320      755321      755322      755323      755324      755325      755326      755327      755328      755329      755330      755331      755332      755333      755334      755335      755336      755337      755338      755339      755340      755341      755342      755343      755344      755345      755346      755347      755348      755349      755350      755351      755352      755353      755354      755355      755356      755357      755358      755359      755360      755361      755362      755363      755364      755365      755366      755367      755368      755369      755370      755371      755372      755373      755374      755375      755376      755377      755378      755379      755380      755381      755382      755383      755384      755385      755386      755387      755388      755389      755390      755391      755392      755393      755394      755395      755396      755397      755398      755399      755400      755401      755402      755403      755404      755405      755406      755407      755408      755409      755410      755411      755412      755413      755414      755415      755416      755417      755418      755419      755420      755421      755422      755423      755424      755425      755426      755427      755428      755429      755430      755431      755432      755433      755434      755435      755436      755437      755438      755439      755440      755441      755442      755443      755444      755445      755446      755447      755448      755449      755450      755451      755452      755453      755454      755455      755456      755457      755458      755459      755460      755461      755462      755463      755464      755465      755466      755467      755468      755469      755470      755471      755472      755473      755474      755475      755476      755477      755478      755479      755480      755481      755482      755483      755484      755485      755486      755487      755488      755489      755490      755491      755492      755493      755494      755495      755496      755497      755498      755499      755500      755501      755502      755503      755504      755505      755506      755507      755508      755509      755510      755511      755512      755513      755514      755515      755516      755517      755518      755519      755520      755521      755522      755523      755524      755525      755526      755527      755528      755529      755530      755531      755532      755533      755534      755535      755536      755537      755538      755539      755540      755541      755542      755543      755544      755545      755546      755547      755548      755549      755550      755551      755552      755553      755554      755555      755556      755557      755558      755559      755560      755561      755562      755563      755564      755565      755566      755567      755568      755569      755570      755571      755572      755573      755574      755575      755576      755577      755578      755579      755580      755581      755582      755583      755584      755585      755586      755587      755588      755589      755590      755591      755592      755593      755594      755595      755596      755597      755598      755599      755600      755601      755602      755603      755604      755605      755606      755607      755608      755609      755610      755611      755612      755613      755614      755615      755616      755617      755618      755619      755620      755621      755622      755623      755624      755625      755626      755627      755628      755629      755630      755631      755632      755633      755634      755635      755636      755637      755638      755639      755640      755641      755642      755643      755644      755645      755646      755647      755648      755649      755650      755651      755652      755653      755654      755655      755656      755657      755658      755659      755660      755661      755662      755663      755664      755665      755666      755667      755668      755669      755670      755671      755672      755673      755674      755675      755676      755677      755678      755679      755680      755681      755682      755683      755684      755685      755686      755687      755688      755689      755690      755691      755692      755693      755694      755695      755696      755697      755698      755699      755700      755701      755702      755703      755704      755705      755706      755707      755708      755709      755710      755711      755712      755713      755714      755715      755716      755717      755718      755719      755720

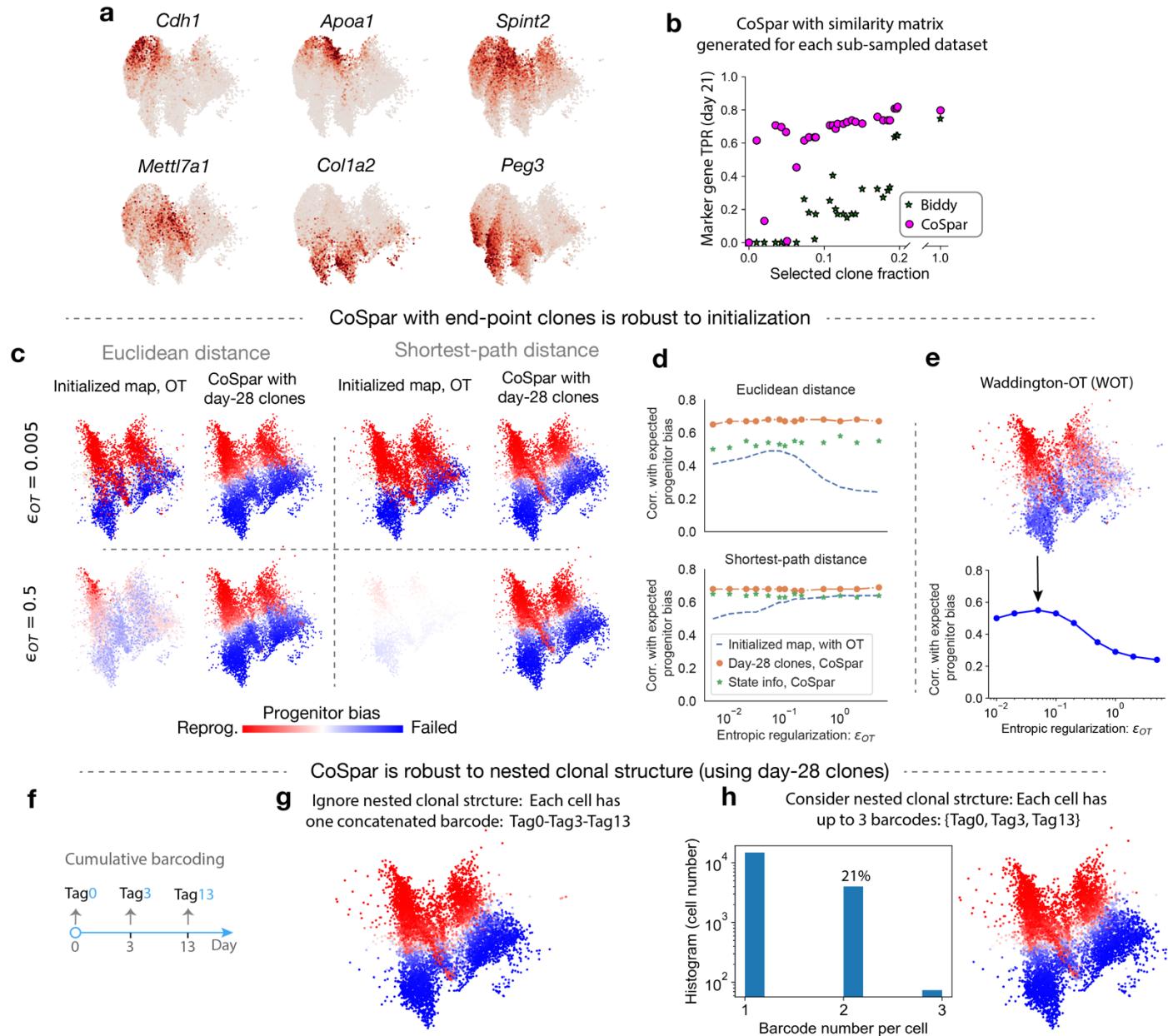


### Supplementary Fig. 3. Benchmarking CoSpar in hematopoiesis.

**a**, CoSpar reconstructs transition maps from sub-sampled and dispersed clonal data. Here, we evaluate the prediction error as the Wasserstein distance between fraction of cell progeny predicted to occupy a given fate, compared to that obtained from the ‘ground truth’ transition map constructed using all clonal data rooted in day 2 clones (see main text). In **a**, the prediction error is assessed for a decreasing fraction of day 4-6 clones, obtained by progressively excluding less dispersed clones that contribute the strongest signal (see Fig. 4**b**). Green curve is obtained by applying the method from the original paper. A lower bound on the error (random split distance) is the Wasserstein distance between random 50% partitions of the ground-truth data. The largest observed errors are

1 comparable to the Wasserstein distance between populations separated by two days of progressive  
2 differentiation (upper grey arrow).  
3 **b**, The ground truth and predicted fate maps for neutrophils cluster using the 15% most dispersed  
4 clones. These plots illustrate one value on the plot in **a**.  
5 **c**, The normalized covariance of clonal barcode abundances between different cell types, calculated  
6 using all data on day 4 of differentiation<sup>1</sup>.  
7 **d**, The correlation of predicted transition probabilities of progenitors, inferred with CoSpar using  
8 different data indicated (See Methods).  
9 **e**, Joint CoSpar optimization is robust to initialization and choice of distance metric. This panel  
0 accompanies Fig. 4g. Plots show the correlation of progenitor biases calculated from the transition  
1 maps for different initialization choices of the transition map. Optimal transport (OT) is used to  
2 initialize the transition map from state information alone prior to CoSpar. Plots scan the OT entropic  
3 regularization strength  $\epsilon_{OT}$ .  
4 **f**, Application of Waddington-OT (WOT) to hematopoiesis dataset. WOT was applied to the same  
5 data in Ref<sup>2</sup>, where clonal data was used to tune the local cell proliferation rates. When WOT is  
6 applied without access to any clonal information, performance is degraded as seen by comparing the  
7 plots here to the ground truth. Plots are to be compared with those in panels **c,d** and Fig. 4c. WOT is  
8 applied with default parameters ( $\epsilon_{OT} = 0.05$ ).  
9 **g-i**, Predicting early fate boundaries in the Gata1+ lineages using the original method from Ref<sup>2</sup>. **g**,  
0 Predicted progenitor bias among the Gata1+ cells on the state embedding. **h**, Comparison of the  
1 number of differentially expressed genes (FDR<0.05) identified from different methods of clonal  
2 analysis. **i**, Gene expression heat map for all differentially expressed genes identified with the  
3 Weinreb method<sup>2</sup>.

4  
5



### Supplementary Fig. 4. Benchmarking CoSpar in fibroblast reprogramming.

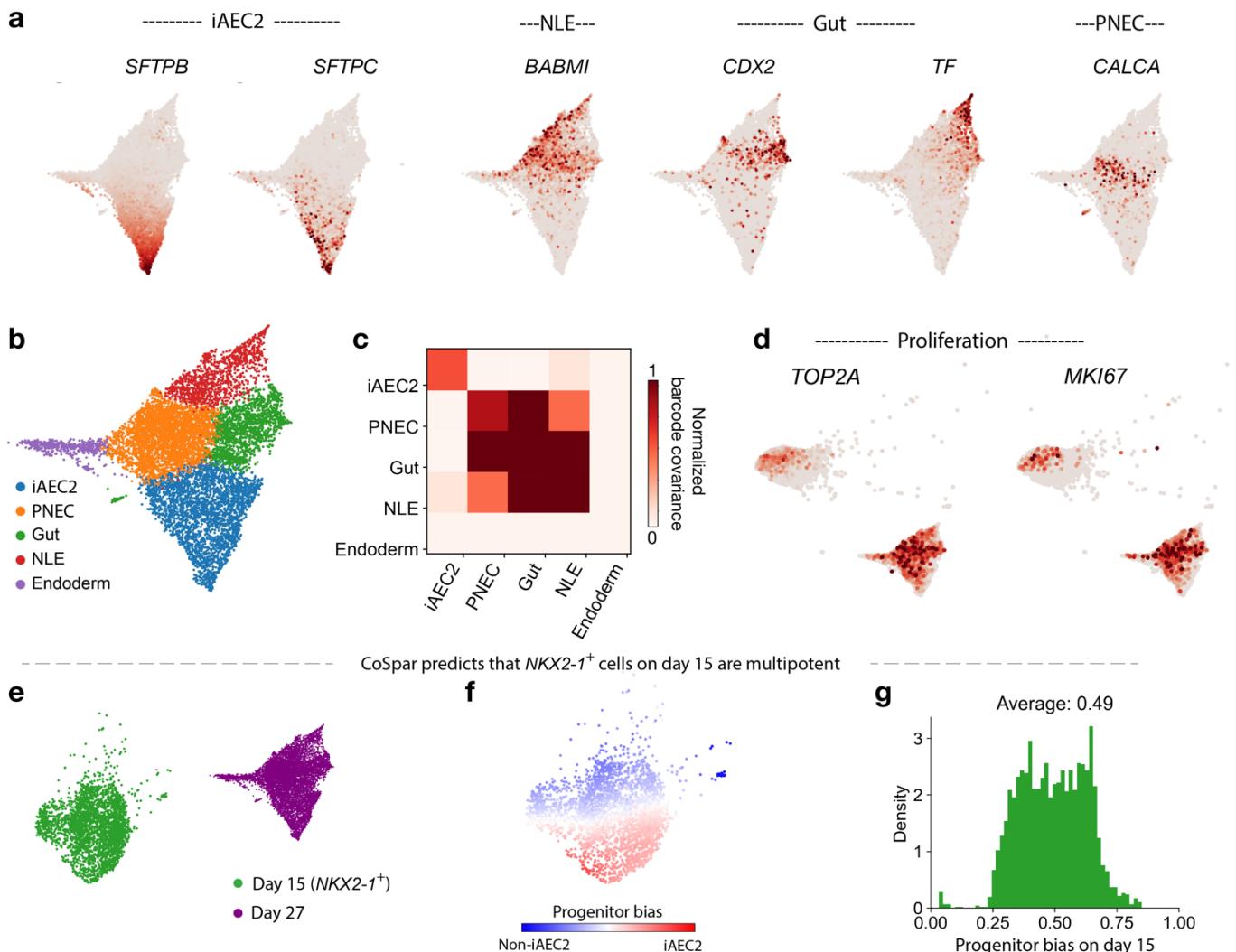
a, Expression of selected marker genes on UMAP visualizations from day 15, 21 and 28.

b, Reproduction of results in Fig. 5e using a similarity matrix obtained from each sub-sampled dataset. Results are seen to be robust to sub-sampling strategies.

c-e, Transition maps inferred by CoSpar with access only to end-point clonal information are robust to the choice of initialization. These panels accompany Fig. 5h. c, Visualization of the progenitor bias derived from the initialized transition map and the corresponding CoSpar prediction, for different entropic regularizations and distance metrics as indicated. d, Parameter sweep quantifying the stability of the predicted progenitor bias. e, Progenitor bias prediction from Waddington-OT<sup>3</sup>, which relies only on state information. Upper panel: the predicted progenitor bias on the state manifold at  $\epsilon_{OT}=0.05$ . Lower panel: progenitor bias correlation with ground truth across different  $\epsilon_{OT}$  values.

f-h, CoSpar analysis with clonal barcodes integrated at sequential time points. The analysis was done with clonal data on day 28. f, The cumulative barcoding scheme in the reprogramming experiment.

1 Cells were barcoded on day 0, 3, and 13. **g**, A progenitor bias prediction generated by concatenating  
 2 all tags from all three time points into a single clonal barcode for each cell, thus ignoring the nested  
 3 clonal structure in the data. **h**, Equivalent results of CoSpar analysis with nested clonal structure,  
 4 carried out by treating Tag0, Tag3 and Tag13 as independent barcodes for a cell, such that each cell  
 5 may have up to three barcodes. Left panel shows the histogram of barcode number per cell.

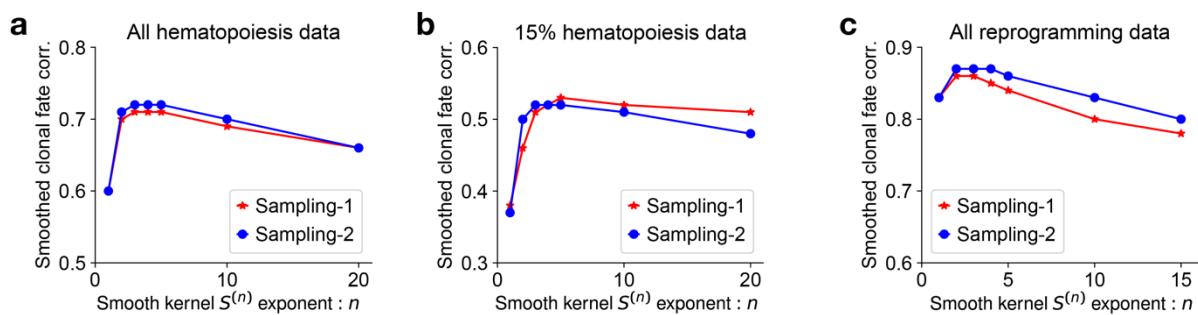


4

5 **Supplementary Fig. 5. Marker gene expression and clonal structure during differentiation into**  
 6 **alveolar cells and other endodermal cells.**

7 **a**, Expression of genes associated (in Ref<sup>4</sup>) with iAEC2 cells, non-lung endoderm (NLE), gut  
 8 endoderm, and pulmonary neuroendocrine cells (PNEC).

1 **b**, Leiden clustering of day-27 cell states. Cluster are named based on their corresponding gene  
 2 expression.  
 3 **c**, Normalized barcode covariance on day 27 among all clusters, showing evidence of clonal  
 4 partitioning of iAEC2 cells.  
 5 **d**, Expression of two representative genes marking proliferating cells (*TOP2A* and *MKI67*) on day 17  
 6 and 27 state manifold, showing that cells predicted by CoSpar to show low commitment on day 17  
 7 appear proliferating (Fig. 6c).  
 8 **e-g**, CoSpar predicts that lineage restriction occurs after day 15, except for a rare fraction of cells  
 9 committed to non-iAEC2 fates. **e**, UMAP visualization of cell states on day 15 and 27. **f**, CoSpar-  
 0 predicted progenitor bias among cells on day 15. **g**, Histogram of the progenitor bias on day 15  
 1 (shown in panel f). Unlike on day 17 (Fig. 6c), here progenitor bias is concentrated at 50%.



7 **Supplementary Fig. 6. Establishing upper bounds for fate prediction after data loss.** In this  
 8 paper, performance of CoSpar was compared to previously published methods by discarding clonal  
 9 data and then examining the fidelity of fate predictions in the face of data loss. Supporting the results  
 0 reported in Figs. 4g,i and 5h, we obtain an upper bound for fate prediction, by randomly sampling  
 1 50% cells from the full ground-truth dataset in each case to predict the progenitor bias of remaining  
 2 cells, with different smoothing exponents  $n$ . Prediction was carried out by first inferring the progenitor  
 3 bias  $Q_i^{tr}$  from the training data (denoted by  $tr$ ) to predict the bias  $Q_i^{tst}$  of the test data, by imputation  
 4 via graph diffusion:  $Q_i^{tst} = \sum_j S_{ij}^{(n)} Q_j^{tr}$ . Results show that, in all the three cases considered, a  
 5 smoothing exponent  $n=3$  provided the best correlation between the imputed and actual values of  $Q_i^{tst}$ .  
 6 These correlation values are indicated by the upper dashed grey lines in Figs. 4g,i and 5h.

2

## CONTENTS

3	Supplementary Note 1: Connecting transition maps to models of differentiation	10
4	Supplementary Note 2: The effect of noisy measurement on transition map inference	11
5	Supplementary Note 3: Coherent sparse optimization	12
6	Supplementary Note 4: Transition map initialization with HighVar	13
7	References	14

8 **Supplementary Note 1: Connecting transition maps to models of differentiation**

9 This note grounds the finite-time transition map in a stochastic model of cell differentiation. In doing so it also  
10 clarifies what cannot be learnt from the transition map.

11 We begin by considering a Markov model of differentiation represented by an arbitrary graph of finite size, where  
12 each node represents a cell state. In this model, each cell probabilistically undergoes proliferation, death, and  
13 differentiation with rates that are specific to the cell state. A clone is a realization of such a stochastic branching  
14 process, seeded as a single barcoded cell in some cell state. Starting from a cell state  $i$ ,  $k_{ij}$  is the differentiation rate  
15 to a different state  $j$ ;  $b_i$  is the probability of a cell dividing into two cells; and  $d_i$  is the cell loss rate for cells in state  
16  $i$ . We assume that these rates are first-order (independent of the number of cells in a state). These rates can vary  
17 with time to reflect changes in the tissue environment. Supplementary Fig. 1a shows a simplified example of such a  
18 model.

20  
21 This model is useful in its simplicity, but it is clearly not general: being a Markov process, it assumes that we have  
22 a complete measurement of the variables that could affect state dynamics, such as the transcriptome, epigenome,  
23 and extracellular environment. This is unlikely to be true. Incomplete state measurement leads to a non-Markovian  
24 dynamics<sup>5</sup>. Nonetheless, our model may be a useful approximation as it generates predictions of biomarkers and fate  
25 regulators, and their correlation with fate bias.

26 Our goal in this paper is to learn the structure of such a graphical model (e.g. Supplementary Fig. 1a) and its rate  
27 constants, from LT-scSeq data. To learn a model from data, we focus most simply on the mean dynamics of cell number  
28 at each state. To do so, one could consider a complete stochastic description using the chemical master equation<sup>6</sup>,  
29 which gives the distribution evolution over the extended state space  $N \times X = \{(N_i, X_i) \ \forall i; \text{ and } N_i = 1, 2, \dots\}$ ,  
30 where  $N_i$  is the number of cells at state  $i$  and  $X_i$  is the corresponding state. However, because we assume a first-order  
31 model, there exists a closed-form equation for the dynamics of average cell number  $\bar{N}_i(t)$  at state  $i$  and time  $t$ ,

$$\frac{d}{dt} \bar{N}_i(t) = \sum_j \bar{N}_j(t) K_{ji}, \quad (1)$$

33 where  $K_{ij} \equiv (1 - \delta_{ij})k_{ij} + \delta_{ij}(b_i - d_i - \sum_{k \neq i} k_{ik})$ , with  $\delta_{ij} = \{1 \text{ if } i = j; \text{ otherwise } 0\}$ , is the instantaneous  
34 transition rate from state  $i$  to  $j$  that includes all cellular processes: division, cell death, and differentiation. This  
35 mean dynamics only captures the net effect of cell number change ( $b_i - d_i$ ), and does not distinguish whether it is  
36 from cell proliferation or loss.

37 To make contact with experiment, we represent the number of cells at each state as a fraction of the total cell  
38 number to obtain the cell density:  
39

$$P_i(t) \equiv \frac{\bar{N}_i(t)}{\bar{N}(t)}, \quad (2)$$

40 where  $\bar{N}(t) \equiv \sum_j \bar{N}_j(t)$  is the total cell number at time  $t$ . The dynamics of the cell density  $P_i(t)$  is

$$\frac{d}{dt} P_i(t) = \sum_j P_j \tilde{K}_{ji}(t), \quad (3)$$

41 where  $\tilde{K}_{ji}(t) \equiv K_{ji} - \delta_{ji}\bar{\alpha}(t)$ , and  $\bar{\alpha}(t) \equiv \sum_k P_k(t)(b_k - d_k)$  is the average growth rate of the population at time  $t$ .  
42 Diagonal elements in  $\tilde{K}$  reflect whether net growth in each state is larger (positive) or smaller (negative) than the  
43 population average.

44 We now can ground the transition map  $T$  in terms of the model. Integrating Eq. (3) from time  $t_1$  to  $t_2$  leads to the  
 45 relation

$$P_i(t_2) = \sum_j P_j(t_1) T_{ji}(t_1, t_2), \quad (4)$$

46 where the intrinsic finite-time transition map

$$T = \exp \left( \int_{t_1}^{t_2} \tilde{K} dt \right) \quad (5)$$

47 is obtained from matrix exponentiation of the corrected instantaneous transition rate matrix  $\tilde{K}$ .

48 The transition probability  $T_{ij}$  is the fraction of progenies from initial state  $i$  that ends at later state  $j$  (Supplementary  
 49 Fig. 1b). To see this, we can sum over all states in Eq. (4), and noting that  $\sum_i P_i(t) = 1$ , we have  $1 = \sum_j P_j(t_1) \sum_i T_{ji}$ .  
 50 This equation is valid for any distribution  $P_j(t_1)$  and therefore the transition map satisfies the conservation property

$$\sum_j T_{ij} = 1. \quad (6)$$

51 Owing to its normalization (Eq. 6), the transition map that is experimentally accessible captures the most interesting  
 52 property we want: the probability of a cell to differentiate into different cell types. A certain initial state  $i$  can transition  
 53 to multiple states over time window  $t$ , i.e.,  $T$  has multiple non-zero entries associated with the  $i$ -th row.

54 Nonetheless, it is important to note that  $T_{ij}$  is shaped both by differences in transition rates between states, and  
 55 by the collective effect of proliferation and cell death along the trajectories between state  $i$  and  $j$ . Mathematically,  
 56 although proliferation and cell death only affect the diagonal terms in the instantaneous transition matrix  $\tilde{K}$ , the  
 57 matrix exponentiation in Eq. (5) will propagate this effect to the off-diagonal terms in the finite-time transition  
 58 matrix  $T$ . For this reason, empirical transition maps alone obscure differences between biases in proliferation and  
 59 choice towards competing fates, as illustrated in Supplementary Fig. 1d.

## 60 Supplementary Note 2: The effect of noisy measurement on transition map inference

61 In Eq. (5), the transition map is seen to emerge from stochastic state transitions accumulating over time. In  
 62 practice, an inferred map is also shaped by sources of noise associated with measurement and subsequent dimen-  
 63 sionality reduction of the data. In this note, we examine the errors propagated from different technical sources into  
 64 the observed transition map  $T$ . As might be expected, we show that technical sources of error lead to a ‘blurred’  
 65 transition map, delocalized over the cell state graph. The smoothing kernels connecting the true and observed transi-  
 66 tion map can be understood as a matrix product of error kernels associated with each individual source of uncertainty.  
 67

a. *Measurement errors.* We will consider the errors associated with correctly assigning transition rates from a state  $i$  at time  $t_1$  to state  $j$  at time  $t_2$ . Such a transition contributes to mass at matrix element  $T_{ij}(t_1, t_2)$  of the transition map. At time  $t_2$ , errors in measurement re-assign cells from state  $j$  to another state  $k$ , with a probability  $\epsilon_{jk}$  normalized such that  $\sum_k \epsilon_{jk} = 1$ . With such an error, the observed transition map now becomes  $T_{ij}^{(\text{obs.})} = \sum_k T_{ik} \epsilon_{kj}$ . A similar error may occur at  $t_1$ . Because technical errors may differ between time points, we will denote  $\epsilon^{(i)}$  as the error in measuring the state of a cell at time  $t_i$ . Accounting for errors in two time points, the observed transition map now becomes:

$$T_{ij}^{(\text{obs.})} = \sum_{k,l} \epsilon_{ki}^{(1)} T_{kl} \epsilon_{lj}^{(2)}.$$

b. *Clonal dispersion.* In LT-scSeq experiments, the cells sampled at  $t_1$  are clonally related to those that give rise to cells sampled at  $t_2$ . But being distinct, they may occupy different states. As above, we consider the error in estimating transition rates from state  $i$  at  $t_1$  to state  $j$  at  $t_2$ . At  $t_1$ , a clonally-related state,  $k$ , is observed instead of state  $i$ , with a probability that we shall denote  $\sigma_{ik}$ . This probability satisfies normalization  $\sum_k \sigma_{ik} = 1$ . Accounting for this clonal dispersion, the observed transition map relates to the true transition map through the relation:

$$T_{ij}^{(\text{obs.})} = \sum_k \sigma_{ki} T_{kj}.$$

68 Note that because cells divide, more than one cell may be observed in a clone at time  $t_1$ . In this case, the error  
 69 kernel  $\sigma_{ki}$  no longer has a unique definition because choices in constructing the transition map may assign more or  
 70 less weight to particular cells within each clone. By enforcing local coherence, CoSpar strongly weights  $\sigma_{ki}$  towards  
 71 states  $k$  that are close to  $i$ , thus reducing errors in the transition map as compared to using a ‘naive’ clonal analysis  
 72 method such as we have previously reported<sup>2</sup>, which weights all cells in a clone at  $t_1$  equally.

Compounding clonal dispersion and measurement error, we recognize the the observed transition map has the form:

$$T^{(\text{obs.})}(t_1, t_2) = S_1^T T(t_1, t_2) S_2,$$

73 where  $S_1 = \epsilon^{(1)} \sigma$  and  $S_2 = \epsilon^{(2)}$ .

#### 74 Supplementary Note 3: Coherent sparse optimization

75 Our goal in dynamic inference is to learn the finite-time transition map, as defined in Eq. (4), for the set of observed  
 76 cell states in a given experiment. After imposing sparsity and coherence constraints (see main text), we obtain the  
 77 cost function,

$$\min_T \|T\|_1 + \alpha \|LT\|_2, \text{ s.t. } \sum_m \|P(t_2; m) - P(t_1; m)T(t_1, t_2)\|_2 \leq \epsilon; T \geq 0; \text{ Normalization.} \quad (7)$$

78 Here,  $P(t_{1,2}; m)$  is a row-vector representing the distributions of cell states within the  $m$ -th clone.  $L_{ij} = 1 -$   
 79  $\bar{w}_{ij} / \sum_j \bar{w}_{ij}$  is the normalized graph laplacian, with  $w_{ij}$  the graph connectivity of the nearest neighbor kNN graph  
 80 of cell states. Defining  $\mathbf{P}(t)$  as a clone-by-cell matrix resulting from concatenation of individual clonal distribution:  
 81  $\{P(t; m), m = 0, 1, 2, \dots\}$ , we note that  $\sum_m \|P(t_2; m) - P(t_1; m)T(t_1, t_2)\|_2 = \|\mathbf{P}(t_2) - \mathbf{P}(t_1)T(t_1, t_2)\|_2$ , which gives the  
 82 form of the cost function given in the main text. For joint optimization, the cost function is additionally minimized  
 83 over  $\mathbf{P}(t_1)$ , i.e.  $\min_{\mathbf{P}(t_1)} [\dots]$ .

84 Before continuing, we note the relationship of this optimization problem to past literature. Absent the coherence  
 85 constraint ( $\alpha = 0$ ), this optimization problem reduces to sparse optimization by lasso regression. To our knowledge,  
 86 only one study has explored the extension of lasso to enforce coherence with relation to a data embedding, called  
 87 ‘fused lasso’ optimization<sup>7</sup>. Fused lasso is however different in three important ways from Eq. (7). First, it suppresses  
 88 the first-order derivative of the inference target to promote coherence. Second, fused lasso was developed for 1-d  
 89 or 2-d datasets, assuming a natural ordering for the observed cell states. Third, like lasso, the inference object of  
 90 fused lasso is a vector. In contrast, the coherent sparse optimization in Eq. (7) is generalized to arbitrary graphs;  
 91 it suppresses the second-order derivative (the curvature) to enforce coherence; and it is generalized to matrix inference.

92 Our goal is now to ground the optimization problem in LT-scSeq data, and to propose an algorithm that approx-  
 93 imates solution of Eq. (7). To make connection with raw clonal data, we approximate the density profile matrices  
 94  $\mathbf{P}(t)$  as,

$$\mathbf{P}(t) = I(t)S(t), \quad (8)$$

96 where  $I(t)$  is a clone-by-cell matrix observed at time  $t$ , and  $S(t)$  is a cell-cell similarity matrix at time  $t$ . Note that  
 97 Eq. (8) integrates the state information (encoded in  $S(t)$ ) and clonal information (encoded in  $I(t)$ ) into  $\mathbf{P}$ . This local  
 98 smoothing operation indirectly imposes coherent transitions in this system.

99 We now discuss implementation of the optimization problem. Eq. (7) might be formulated as a quadratic program-  
 100 ming problem, and be solved accordingly as in fussed lasso<sup>7</sup>. However, this strategy is very expensive computationally<sup>7</sup>.  
 101 There could be ways to solve the optimization efficiently and exactly, and we leave it as an open problem. Instead,  
 102 we provide an efficient yet heuristic way to solve the optimization. Specifically, we break down individual elements of  
 103 the objective function, and propose a simple alternative for each of them.

105 1. *Sparsification*. Instead of including the sparsity term  $\|T\|_1$  into the objective function, we directly apply a  
 106 pre-defined thresholding to the transition map at each iteration:  $T \leftarrow \theta(T, \nu)$ , where

$$[\theta(T, \nu)]_{ij} = \begin{cases} T_{ij}, & \text{if } T_{ij} \geq \nu \max_j T_{ij} \\ 0, & \text{Otherwise} \end{cases} \quad (9)$$

107 2. *Transitions within clones.* To enforce Eq. (4) for each observed clone, we consider a clonal transition map  $\pi^m$   
 108 for the  $m$ -th clone, which allows only intra-clone transitions and conserves the total transition flux within a  
 109 clone. We do so by projecting the transition map  $T$  and performing clone-wise normalization:  $\pi^m \leftarrow \mathcal{P}_m(T)$ :

$$[\mathcal{P}_m(T)]_{ij} = \frac{\tilde{\pi}_{ij}^m}{\sum_{i'j'} \tilde{\pi}_{i'j'}^m}, \quad (10)$$

110 where  $\tilde{\pi}_{ij}^m = T_{ij}$  if the transition  $i \rightarrow j$  occurs within clone  $m$ , and otherwise  $\tilde{\pi}_{ij}^m = 0$ . The composite map  
 111 capturing all intra-clone transitions is then,

$$\mathcal{P}(T) = \sum_m \mathcal{P}_m(T) \quad (11)$$

112 A map constructed in this way,  $\pi = \mathcal{P}(T)$ , will satisfy the following equation approximately:

$$I(t_2) \approx I(t_1)\pi(t_1; t_2), \quad (12)$$

113 which is the clonal constraint for directly observed cell states<sup>8</sup>. The map  $\pi(t_1; t_2)$  can be used to specify  $T$ , but  
 114 being constrained to clones it is no longer coherent.

115 3. *Coherence.* To enforce coherence, we begin by noting that Eqs. (4), (8) and (12) together lead to the relationship  
 116  $T(t_1; t_2) = S_{t_1}^{-1}\pi(t_1; t_2)S_{t_2}$ . As similarity matrices  $S$  are generally non-invertable, we introduce a pseudo-inverse,

$$T(t_1; t_2) \approx S_{t_1}^+ \pi(t_1; t_2) S_{t_2}. \quad (13)$$

117 Eq. (13) smoothes the transition map learnt within-clones,  $\pi$ , over nearby states to get a transition map  $T$   
 118 across all states.  $T$  is now a locally continuous map and satisfies the coherence constraint: similar initial cell  
 119 states have similar fate outcomes.

120 This approach to calculating  $T$  leads to minimization of the term  $\alpha||LT||_2$  in Eq. (7), although the parameter  
 121  $\alpha$  establishing the relative weight of coherence is no longer explicitly identifiable in the procedure. It is instead  
 122 reflected in the extent of smoothing.

123 These three steps, carried out sequentially and iteratively, define the CoSpar algorithm given in methods. Note that  
 124 normalization is performed clone-wise in Eq. (11). The non-negativity constraint,  $T \geq 0$ , is implicitly satisfied in the  
 125 above steps. In our strategy, Eq. (13) is the most time-consuming step as it involves multiplication of large matrices.  
 126 CoSpar is nonetheless efficient as it carries out matrix multiplication *only* at Eq. (13), and we find that it converges  
 127 within a few iterations (Supplementary Fig. 2d).

#### 128 Supplementary Note 4: Transition map initialization with HighVar

129 The HighVar method provides an approach to initialize the joint optimization of  $T$  and  $I(t_1)$  (see Methods). The  
 130 approach is loosely motivated by the expectation that cells similar in gene expression between time points may share  
 131 clonal origin. This expectation can be violated; we use it only to initialize numerical optimization.

132 HighVar consists of three steps: 1) Select highly variable genes that are expressed at both  $t_1$  and  $t_2$ ; 2) For each  
 133 highly variable gene (indexed by  $m$ ), threshold its expression to form a binary expression matrix  $\hat{x}_{im} \in \{0, 1\}$  for all  
 134 states observed at  $t_1$  and  $t_2$  to generate pseudo clonal data  $\hat{I}(t_1)$  and  $\hat{I}(t_2)$  from the binary expression matrix; 3) Run  
 135 CoSpar with  $\hat{I}(t_1)$  and  $\hat{I}(t_2)$ . The pseudo-clonal data  $\hat{I}(t_1)$  and  $\hat{I}(t_2)$  are discarded, and the resulting map  $T$  is used  
 136 to initialize CoSpar with the true clonal data.

137 For the first step, we use the SPRING gene filtering function filter\_genes with an adjustable gene variability percentile  
 138 parameter HighVar\_gene\_pctl to select highly variable genes<sup>9</sup>. For the second step we discretize the gene expression  
 139 of each highly-variable gene, sequentially, with a gene-specific threshold  $\eta_m$ :

$$\hat{I}_{im} = H(x_i(m) - \eta_m) \times Z_{im},$$

140 where  $H(\cdot)$  is the Heaviside step function ( $H(x) = 1$  if  $x > 0$ ; otherwise 0),  $Z_{im} = [1 - H(\sum_{m^*=0}^{m-1} \hat{I}_{im^*})]$  sums  
 141 over previously considered genes to ensure that the same cell is not assigned to more than one pseudo-clone. The  
 142 gene-specific threshold  $\eta_m$  is chosen such that every pseudo clone has the same number of cells at each time point  
 143  $N_t/M$ , where  $N_t$  is the number of observed cells at time  $t$  and  $M$  is the total number of highly variable genes (i.e.,

<sup>144</sup> pseudo clones). In case  $N_t/M$  is not an integer, we use its ceil, i.e.,  $\lceil N_t/M \rceil$ , and stop the clonal matrix update when  
<sup>145</sup> all cells are clonally labeled.

---

<sup>146</sup> [1] C. Weinreb and A. M. Klein, Proceedings of the National Academy of Sciences **117**, 17041 (2020).  
<sup>147</sup> [2] C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, and A. M. Klein, Science **367** (2020).  
<sup>148</sup> [3] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, *et al.*, Cell  
<sup>149</sup> **176**, 928 (2019).  
<sup>150</sup> [4] K. Hurley, J. Ding, C. Villacorta-Martin, M. J. Herriges, A. Jacob, M. Vedaie, K. D. Alysandratos, Y. L. Sun, C. Lin, R. B.  
<sup>151</sup> Werder, *et al.*, Cell Stem Cell **26**, 593 (2020).  
<sup>152</sup> [5] S.-W. Wang, K. Kawaguchi, S.-i. Sasa, and L.-H. Tang, Phys. Rev. Lett. **117**, 070601 (2016).  
<sup>153</sup> [6] D. T. Gillespie, The journal of physical chemistry **81**, 2340 (1977).  
<sup>154</sup> [7] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, Journal of the Royal Statistical Society: Series B (Statistical  
<sup>155</sup> Methodology) **67**, 91 (2005).  
<sup>156</sup> [8] One can appreciate that this equation is approximately satisfied because  $I(t_1)\pi(t_1; t_2)$  gives a matrix with non-zero values  
<sup>157</sup> on at clonally observed states at  $t_2$ . Therefore  $I(t_1)\pi(t_1; t_2)$  has the same sparse structure as  $I(t_2)$  but will differ in the  
<sup>158</sup> exact non-zero values because  $I(t_2)$  is strictly binary.  
<sup>159</sup> [9] C. Weinreb, S. Wolock, and A. M. Klein, Bioinformatics **34**, 1246 (2018).