

LETTER

doi:10.1038/nature20777

Synthetic recording and *in situ* readout of lineage information in single cells

Kirsten L. Frieda, James M. Linton, Sahand Hormoz, Joonhyuk Choi, Ke-Huan K. Chow, Zakary S. Singer, Mark W. Budde, Michael B. Elowitz & Long Cai

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. *Nature* is providing this early version of the typeset paper as a service to our customers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Cite this article as: Frieda, K. L. *et al.* Synthetic recording and *in situ* readout of lineage information in single cells. *Nature* <http://dx.doi.org/10.1038/nature20777> (2016).

Received 30 June; accepted 11 November 2016.

Accelerated Article Preview Published online 21 November 2016.

Synthetic recording and *in situ* readout of lineage information in single cells

Kirsten L. Frieda^{1*}, James M. Linton^{1*}, Sahand Hormoz^{1*}, Joonhyuk Choi³, Ke-Huan K. Chow¹, Zakary S. Singer¹, Mark W. Budde¹, Michael B. Elowitz^{1,2} & Long Cai³

Reconstructing the lineage relationships and dynamic event histories of individual cells within their native spatial context is a long-standing challenge in biology. Many biological processes of interest occur in optically opaque or physically inaccessible contexts, necessitating approaches other than direct imaging. Here, we describe a new synthetic system that enables cells to record lineage information and event histories in the genome in a format that can be subsequently read out in single cells *in situ*. This system, termed Memory by Engineered Mutagenesis with Optical *In situ* Readout (MEMOIR), is based on a set of barcoded recording elements termed scratchpads. The state of a given scratchpad can be irreversibly altered by Cas9-based targeted mutagenesis, and read out in single cells through multiplexed single-molecule RNA fluorescence hybridization (smFISH). To demonstrate a proof of principle of MEMOIR, we engineered mouse embryonic stem (ES) cells to contain multiple scratchpads and other recording components. In these cells, scratchpads were altered in a progressive and stochastic fashion as cells proliferated. Analysis of the final states of scratchpads in single cells *in situ* enabled reconstruction of the lineage trees of cell colonies. Combining analysis of endogenous gene expression with lineage reconstruction in the same cells further allowed inference of the dynamic rates at which ES cells switch between two gene expression states. Finally, using simulations, we showed how parallel MEMOIR systems operating in the same cell can enable recording and readout of dynamic cellular event histories. MEMOIR thus provides a versatile platform for information recording and *in situ*, single cell readout across diverse biological systems.

Somatic mutations occur stochastically and independently in different cells, and are inherited from one cell generation to the next. They can therefore leave a record of lineage relationships, or other information, in the genomes of related cells. Pioneering work showed that sequencing can be used to identify somatic mutations and thereby recover lineage information^{1–6}. However, sequencing has generally required disrupting the spatial context of cells, and somatic mutations are distributed throughout the genome, hindering their identification and analysis. Two recent advances together enable an alternative approach: First, CRISPR/Cas9^{7–9} can target mutagenesis to specific genomic elements, facilitating the continuous and controlled generation of stochastic genetic variation at designated genomic regions. Second, *in situ* single cell analysis by sequential smFISH^{10,11} (seqFISH) allows genetic information to be directly interrogated in a highly multiplexed fashion in individual cells within native tissue. Together, these techniques could in principle permit recording and *in situ* readout of genetic changes at specific loci for lineage reconstruction and event recording.

To implement such a system, we devised a bipartite genetic recording element termed the ‘barcoded scratchpad,’ whose state can be

stochastically altered in live cells and read out *in situ* in single cells by smFISH (Fig. 1a, Extended Data Fig. 1a). The scratchpad element consists of 10 repeat units (20 stem loops¹²). gRNA targeting of Cas9 to the scratchpad generates double-strand breaks that result in its deletion, or ‘collapse.’ (Fig. 1a, b). Adjacent to each scratchpad, we incorporated a co-transcribed barcode (Supplementary Table 1). The barcode and scratchpad components can each be identified using specific sets of smFISH probes (Supplementary Table 2), and thus serve as an addressable ‘bit.’

Using a pool of such barcoded scratchpads enables lineage recording and readout through a two-step process: During cell proliferation, Cas9 generates gradual and stochastic accumulation of collapsed scratchpads in each cell lineage. Subsequently, cells can be fixed and analyzed by seqFISH to identify barcodes and assess their states based on the presence or absence of a colocalized scratchpad signal (Fig. 1c).

To implement the MEMOIR system, we engineered a stable mouse ES cell line, designated MEM-01, incorporating barcoded scratchpads, Cas9, and a scratchpad-targeting gRNA (Fig. 1b). First, we used *piggyBac* transposition¹³ to integrate a set of 28 barcoded scratchpad elements into the genome. We identified a clone in which 13 different barcodes were highly expressed (Extended Data Fig. 1b–d). Within this line, we stably integrated a Cas9 variant containing an inducible degron to allow external modulation of Cas9 activity¹⁴. Finally, we engineered a scratchpad-targeting gRNA expressed from a Wnt-regulated promoter¹⁵ (Methods), to enable both external control as well as recording of Wnt pathway activity.

Using this cell line, we verified that smFISH could detect scratchpad collapse. After 48 hours of Cas9 and gRNA induction, we observed a substantial loss of scratchpad smFISH signal, but not barcode signal (Fig. 2a, b, Extended Data Fig. 2). In contrast, in cells in which MEMOIR recording was not induced, colocalization between barcode and scratchpad signals was observed in ~90% of the transcripts, consistent with expected smFISH accuracies^{16,17} (Fig. 2b, c). Although individual barcoded scratchpad transcripts appeared either collapsed or uncollapsed based on colocalization, cells typically exhibited a mixture of collapsed and uncollapsed scratchpads with the same barcode due to the existence of multiple genomic integrations undergoing independent collapse events (Extended Data Fig. 1b). Together, these results indicate that scratchpad states can be altered and that the fraction of collapsed scratchpads for each barcode can be subsequently read out *in situ*.

The fraction of collapsed scratchpads increased progressively over time after Cas9 and gRNA induction, as required for MEMOIR operation. We observed ~27% decrease in mean colocalization fraction after 48 hours of Cas9 and gRNA induction (Fig. 2b, c). Additionally, the collapse rate correlated with the level of gRNA expression, enabling tuning of collapse rate (Extended Data Fig. 2d). By contrast, in the absence of induction, scratchpad states remained stable (Extended Data Fig. 2e–g). Further, a Cre-activated gRNA functioned similarly to

¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125. ²Howard Hughes Medical Institute. ³Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

*Co-first author

the Wnt-activated gRNA (Extended Data Fig. 3a–d), and scratchpad collapse also occurred in CHO-K1 cells and budding yeast (Extended Data Fig. 3e, f), suggesting that the system design is general. Finally, we verified that seqFISH could enable readout of 13 distinct barcoded scratchpads in single cells using 7 rounds of hybridization (Fig. 2d, e; Methods).

To analyze cell lineage, we activated MEMOIR and allowed cells to grow for 3–4 generations, while performing time-lapse imaging to establish an independent “ground truth” lineage for later validation (Fig. 3a). We then fixed the cells and analyzed their barcoded scratchpads by seqFISH (Fig. 3b). Altogether, we analyzed 108 colonies, including 836 cells.

Inspection of scratchpad collapse patterns revealed lineage information. For example, in one colony, barcode 9 was differentially collapsed between two 4-cell clades, consistent with a collapse event occurring after the first cell division (Fig. 3c, left). Similarly, barcode 2 revealed distinct collapse frequencies between first cousins, but similar frequencies between sister cell pairs (Fig. 3c, middle). Barcode 10 provided additional lineage information, as different sister cell pairs showed collapse frequencies that were similar to each other but different from their cousins (Fig. 3c, right). This example, along with others (Extended Data Figs. 4 and 5), shows how scratchpad collapse patterns can provide insight into lineage relationships.

To analyze lineage reconstruction more systematically, we tabulated scratchpad collapse frequencies for all probed barcodes in each colony (Fig. 3d) and used these data to calculate a cell-to-cell “distance” matrix, representing differences in collapse patterns between each pair of cells (Fig. 3e; Supplementary Information). We then applied a binary hierarchical clustering algorithm based on phylogenetic analysis to these distance scores in order to reconstruct the lineage tree^{18,19} (Fig. 3f; see Methods). Finally, as validation, we compared each reconstructed tree to the actual colony lineage obtained directly from the corresponding time-lapse movie (Fig. 3a).

Across all 108 colonies, we observed a broad distribution of reconstruction fidelity (Fig. 3g, “all colonies”). However, using a bootstrap procedure to rank colonies based on the robustness of reconstruction to resampling of the underlying data, it was possible to identify colonies with more informative scratchpad collapse patterns that therefore reconstructed with higher accuracy (Extended Data Fig. 6; Methods). For example, within the top 20% of colonies ranked by bootstrap, 72% of lineage relationships were correctly reconstructed (Fig. 3g, subset 1 and Extended Data Fig. 6a).

To compare these results to theoretical expectations, we simulated idealized MEMOIR operation in 3-generation binary trees (Methods). As expected, mean reconstruction fidelity increased with the number of distinct scratchpads and required relatively few scratchpads to reach high fidelity. For example, fidelity was $82 \pm 19\%$ for 10 scratchpads and $94 \pm 13\%$ for 20 scratchpads at the experimentally measured collapse rate of ~ 0.1 per scratchpad per cell generation (Fig. 3h). With ~ 8 scratchpads, the performance of these idealized simulations matched that of the bootstrap selected colonies (Fig. 3g, Extended Data Fig. 6b, subset 1), consistent with observations that the majority of the 13 barcoded scratchpads targeted by seqFISH provide robust information. The diversity of states generated corresponds to $\sim 2^8 = 256$ scratchpad configurations, comparable to the number of distinguishable alleles observed by sequencing-based approaches²⁰.

The current implementation of MEMOIR exhibited limited reconstruction depth and accuracy. To better understand the relevant sources of error, we performed more detailed simulations, incorporating empirical measurements of noise in both recording (Cas9 and gRNA expression) and readout (e.g., scratchpad expression and smFISH detection) (Extended Data Fig. 7). Interestingly, stochasticity in Cas9 and gRNA expression, as well as smFISH detection, contributed relatively minor errors in reconstruction. Rather, for a given number of scratchpads, the primary sources of error in reconstruction were stochastic fluctuations in scratchpad

expression, and ambiguities introduced due to multiple incorporations of the same barcoded scratchpad (Fig. 3i, Extended Data Fig. 7; Supplementary Information). Based on this analysis, future versions of MEMOIR can be improved by increasing the number of unique scratchpad variants and reducing noise in their expression (see Supplementary Information for further discussion of potential improvements). These improvements should enable MEMOIR to reconstruct deeper and/or more sparsely sampled trees (Extended Data Figs. 8, 9).

Because MEMOIR is compatible with same-cell measurements of endogenous gene expression through additional rounds of smFISH, it can provide both lineage and endpoint cell state information for the same colony. This combination can provide insight into the dynamics of switching between gene expression states (Fig. 4a). For example, ES cells stochastically transition among states with distinct expression levels of the pluripotency regulator *Esrrb*^{21,22}. To infer the rates of these transitions, we measured *Esrrb* expression, and assigned each cell a probability of being in a high or low *Esrrb* expression state^{23,24} (Fig. 4b; Supplementary Information). Using the MEMOIR-inferred lineage, we found that sisters or first cousins were significantly more likely to appear in the same *Esrrb* expression state compared with pairs of second cousins ($p < 0.004$) (Fig. 4c, d). Using a dynamic inference framework^{24,25}, we further inferred the quantitative rates of switching between states (Fig. 4d, right panel, and Extended Data Fig. 10; Supplementary Information), and verified that they were consistent with direct measurements of switching dynamics²³. Going forward, multiplexed *in situ* transcriptional profiling of endogenous genes^{10,11}, together with MEMOIR, should enable analysis of more complex dynamic cell state transition processes.

The design of MEMOIR provides a platform that can record and read out histories of dynamic cellular events beyond lineage information (Fig. 4e, f). Specifically, orthogonal gRNAs expressed from signal-specific promoters can record multiple intracellular signals onto distinct sets of scratchpads. We simulated binary trees of 6 generations in which different cell lineages experienced distinct time-courses of two input signals (Fig. 4g). In these simulations, one gRNA variant was constitutively expressed solely to enable lineage reconstruction using one set of scratchpads. In addition, each of the signals activated expression of a corresponding gRNA variant, generating collapse events in its own specific set of 50 scratchpads, at a rate proportional to the signal magnitude. By analyzing endpoint scratchpad collapse patterns for all 3 sets of scratchpads, we were able to reconstruct both lineage trees and event histories (Fig. 4e–g; Methods). This reconstruction process takes advantage of the reconstructed lineage tree to map the most likely assignment of collapse events from the signal-recording gRNAs to specific positions on the lineage tree, with a maximum possible time resolution of one cell cycle (since the sequence of collapse events within a cell cycle cannot be distinguished). Thus, over timescales of multiple cell cycles, MEMOIR can enable analysis of the sequence, duration, and magnitude of signals along individual cell lineages (Fig. 4g).

Using genomic DNA as a writable and readable recording medium within living cells is a long-standing goal of synthetic biology^{26–30}. A key application for this technology is to enable analysis of lineage and molecular event histories that unfold in complex and optically inaccessible developmental systems over timescales of multiple cell generations. MEMOIR provides a proof of principle demonstration showing recording and readout of such information with endpoint single cell *in situ* measurements. Importantly, the capacity of MEMOIR can be dramatically extended beyond the current demonstration using more scratchpads with improved designs and highly multiplexed seqFISH^{10,11}. Thus, we anticipate this approach will open up new ways of studying developmental trajectories in developing embryos, tumors, and other systems, eventually enabling us to read, within their native spatial contexts, each cell's own individual “memoir”.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 June; accepted 11 November 2016.

Published online 21 November 2016.

1. Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. & Shapiro, E. Genomic variability within an organism exposes its cell lineage tree. *PLoS Comput. Biol.* **1**, e50 (2005).
2. Salipante, S. J. & Horwitz, M. S. Phylogenetic fate mapping. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5448–5453 (2006).
3. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
4. Wasserstrom, A. *et al.* Reconstruction of cell lineage trees in mice. *PLoS One* **3**, e1939 (2008).
5. Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
6. Evrony, G. D. *et al.* Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
7. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
8. Mali, P. *et al.* RNA-Guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
9. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
10. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
11. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
12. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475–478 (2011).
13. Ding, S. *et al.* Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).
14. Banaszynski, L. A., Chen, L.-C., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995–1004 (2006).
15. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish prickles, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680–685 (2003).
16. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
17. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012).
18. Chapal-Irani, N. *et al.* Comparing algorithms that reconstruct cell lineage trees utilizing information on microsatellite mutations. *PLoS Comput. Biol.* **9**, e1003297 (2013).
19. Sokal, R. R. & Michener, C. D. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* **28**, 1409–1438 (1958).
20. McKenna, A. *et al.* Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* 10.1126/science.aaf7907 (2016).
21. van den Berg, D. L. *et al.* Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol. Cell. Biol.* **28**, 5986–5995 (2008).
22. Kumar, R. M. *et al.* Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516**, 56–61 (2014).
23. Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319–331 (2014).
24. Hormoz, S. *et al.* Inferring cell state transition dynamics from lineage trees and endpoint single-cell measurements. *Cell Systems* doi:10.1016/j.cels.2016.10.015 (2016).
25. Hormoz, S., Desprat, N. & Shraiman, B. Inferring epigenetic dynamics from kin correlations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E2281–E2289 (2015).
26. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8884–8889 (2012).
27. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016).
28. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
29. Hsiao, V., Hori, Y., Rothmund, P. W. K. & Murray, R. M. A population-based temporal logic gate for timing and recording chemical events. *Mol. Syst. Biol.* **12**, 869 (2016).
30. Farzadfard, F. & Lu, T. K. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Martin Budd and Hanqing Li for helpful suggestions. We thank Roy Kishony, and members of the Elowitz and Cai labs for discussions and comments on the manuscript. This research was supported by the Allen Distinguished Investigator Program, through The Paul G. Allen Frontiers Group, NIH R01HD075605 and K99GM118910 (to S.H.), the Gordon and Betty Moore Foundation Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative, and the Beckman Institute pilot program.

Author Contributions K.L.F. and J.M.L. performed the experiments with assistance from S.H., J.C., K.K.C., Z.S.S.; K.L.F. and S.H. analyzed the data; S.H. performed the simulations; M.B.E. and L.C. supervised the project. All authors wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.C. (lcai@caltech.edu) or M.B.E. (melowitz@caltech.edu).

Reviewer Information *Nature* thanks A. Raj and the other anonymous reviewer(s) for their contribution to the peer review of this work.

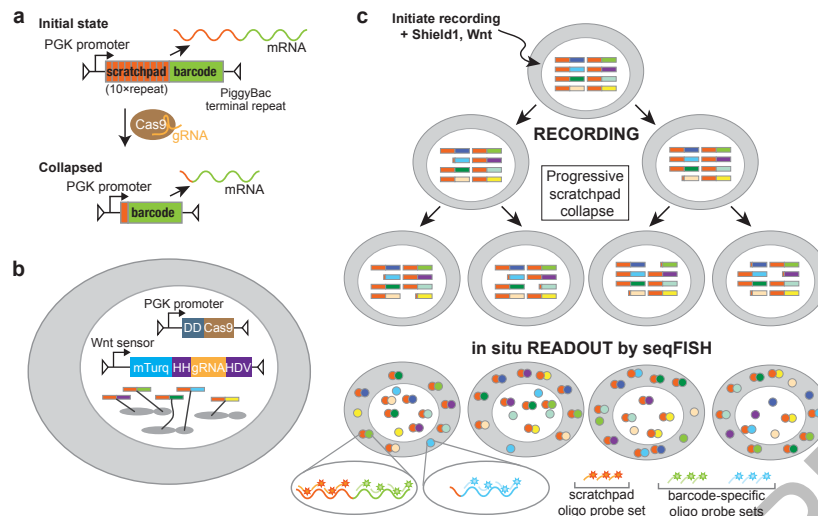


Figure 1 | The MEMOIR system for recording and *in situ* readout of cell lineage. (a) Barcoded scratchpads provide a general purpose recording element whose state can be irreversibly altered by Cas9/gRNA-mediated cleavage. (b) The MEMOIR recording system consists of three types of components, all stably integrated into the genome: (1) a Cas9 variant containing an inducible degenon (DD); (2) a Wnt-inducible gRNA targeting the scratchpad, co-expressed with a fluorescent protein (mTurquoise). Ribozyme sequences (HH, HDV) enable gRNA excision; (3) a set of barcoded scratchpads (2-color elements) integrated throughout the genome. Inverted triangles in (a) and (b) denote piggyBac terminal

repeats, used for genome integration. (c) The MEMOIR recording and readout process: During recording, scratchpads collapse stochastically as cells proliferate, producing distinct scratchpad states in each cell. During readout, individual mRNA molecules are detected with a single scratchpad-specific probe set (orange, inset), and multiple barcode-specific probe sets (blue, green, inset) through sequential rounds of hybridization and imaging. Uncollapsed scratchpads produce colocalized barcode and scratchpad signals (overlapping dots), while collapsed scratchpads produce only a barcode-specific signal (single dots).

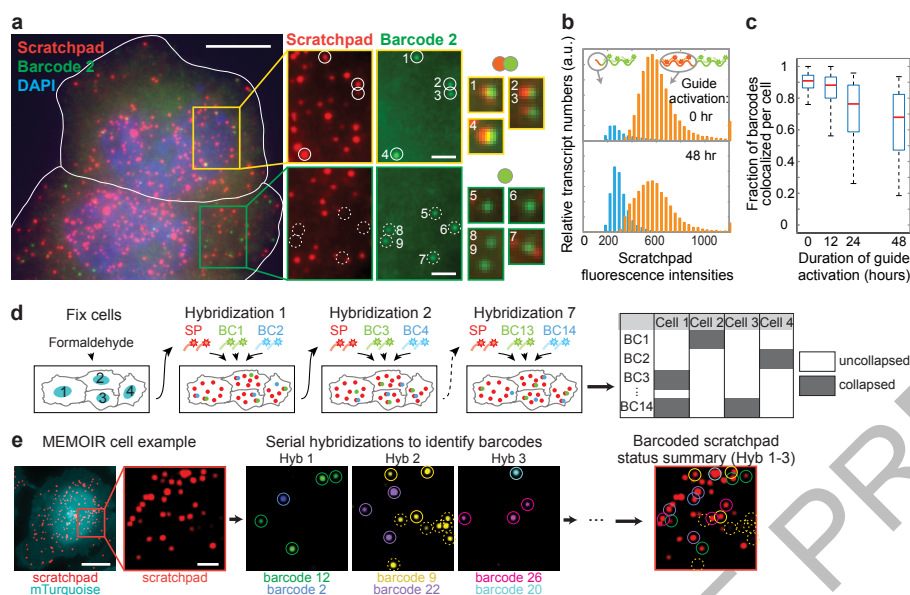


Figure 2 | *In situ* readout of scratchpad state. (a) smFISH readout of scratchpad state in two cells (white outlines). The scratchpad associated with barcode #2 has collapsed in the lower cell, but remains uncollapsed in the upper cell. Overlaid images are slightly offset for visual clarity. (b) Histograms of scratchpad smFISH signal intensities, identified as collapsed (blue) or uncollapsed (orange) based on scratchpad-barcode colocalization. The fraction of collapsed scratchpads increased after 48 hr of activation (upper vs. lower panels). (Far right bars indicate smFISH signal exceeding the maximum displayed intensity.) (c) Scratchpad collapse accumulates over time post-activation. Boxplots show median (red bar), 1st and 3rd quartiles (box) and extrema for four highly expressed barcodes; $n = 1826, 418, 610, 545$ cells, left to right. Activated samples

in (b) and (c) only include gRNA-expressing cells, as measured by co-expressed mTurquoise. (d) Multiplexed readout of barcoded scratchpads (scratchpad, SP; barcode, BC) by sequential rounds of hybridization with distinct probe sets (colors) provide information about the collapse status of multiple barcoded scratchpads in each cell (right). (e) Example of seqFISH analysis. Scratchpads (red) and three pairs of barcodes (middle images) are shown (pseudocolored). Solid and dashed circles at barcode positions indicate uncollapsed and collapsed scratchpads, respectively. Barcode data are superimposed on the scratchpad image in the final panel. For clarity, additional hybridizations and barcodes are not shown. (a,e) Scale bar, $10\mu\text{m}$ (left images) and $2\mu\text{m}$ (zoom-in panels).

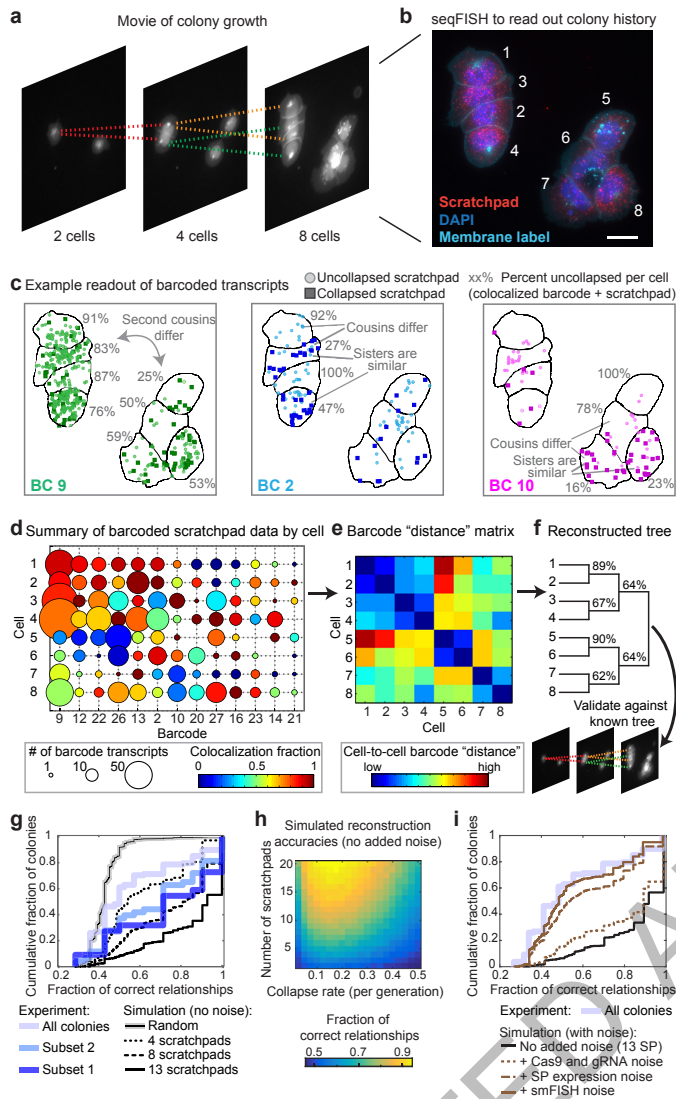


Figure 3 | MEMOIR enables lineage reconstruction in ES cell colonies.

(a) Time-lapse movies of colony growth were acquired to provide lineage "ground truth" (dashed lines) for later validation of reconstructed lineages, but not for reconstruction itself. (b) At the end of the movie, seqFISH was performed, as in Fig. 2. Scale bar, 20 μm . (c) Examples of how barcoded scratchpad collapse patterns reflect cell lineage. (d) MEMOIR readout for the colony in a–c, showing the number of barcode transcripts detected (bubble size) and the uncollapsed fraction (color scale). (e) Data from (d) were used to compute a matrix of cell-cell barcode "distance" (dissimilarity) scores. (f) Reconstructed lineage tree for the same colony (Methods). Percentages on the tree represent frequencies of clade occurrence from a barcode resampling bootstrap procedure. In this case, the reconstructed tree matches that obtained from the movie. (g) Cumulative distributions show the fraction of all pairwise relationships correctly identified in each colony, for all colonies, and for the top 20% (subset 1) or 40% (subset 2) ranked by bootstrap score. (h) Idealized simulations of 3-generation binary trees show how reconstruction accuracy (fraction of relationships correctly identified, color) depends on collapse rate and number of scratchpads. (i) Cumulative distributions from simulations of MEMOIR show how empirically measured noise sources affect reconstruction accuracy in simulated trees, assuming 13 scratchpads (SP). gRNA and Cas9 expression noise adds some reconstruction error (dotted line), which is strongly increased by additional noise from scratchpad expression variability, assuming 2 expressed integrations per barcode (dashed line), and increased slightly more by addition of smFISH readout noise (solid line).

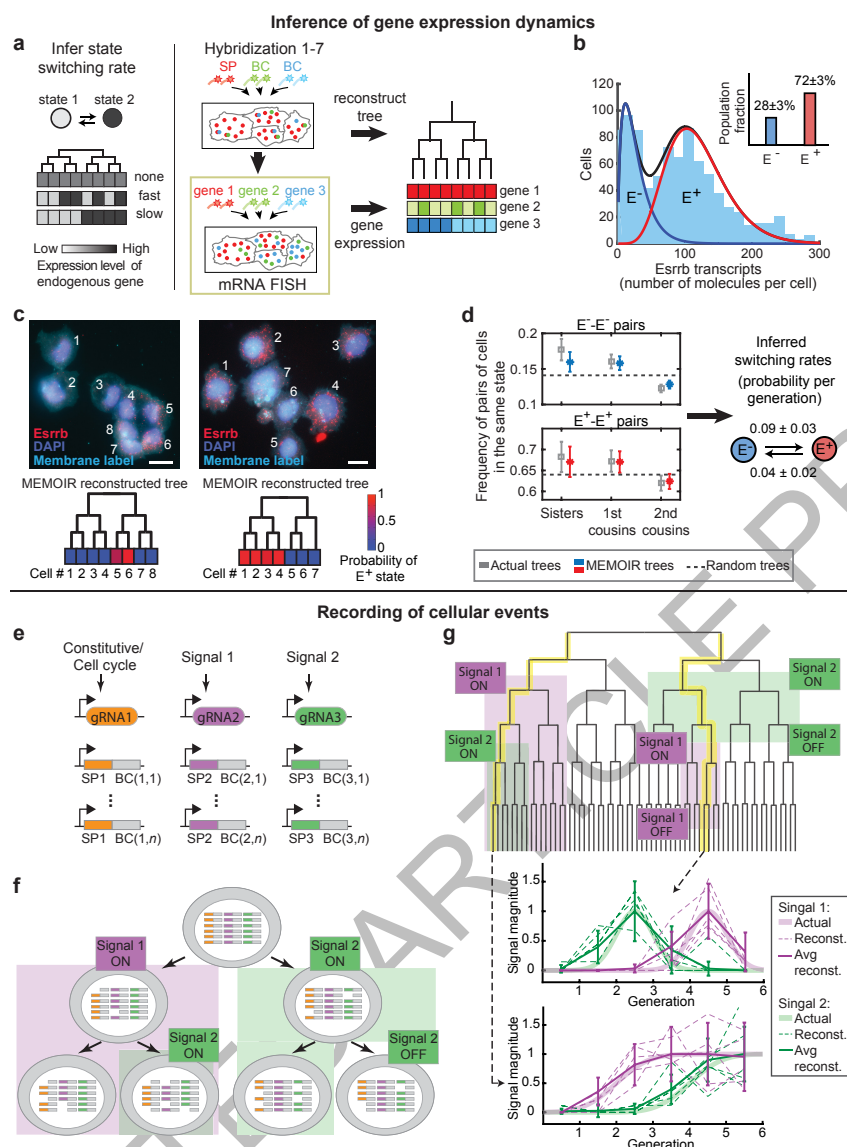


Figure 4 | MEMOIR enables inference of gene expression dynamics and the recording of cellular events. (a–d) Gene expression dynamics inference (see Supplementary Information). (a) The rates of switching between two gene expression states can be inferred by combining reconstructed lineage information and endpoint gene expression measurements (schematic). Inference works because switching rates affect the degree of cell state clustering on endpoints of lineage trees^{24,25}. This analysis can be performed for multiple genes (red, green, blue), which could exhibit different dynamics, as shown schematically. (b) Fits to the bimodal distribution of single cell *Esrrb* transcript counts enable probabilistic assignment of cells to either the low (E^-) or high (E^+) *Esrrb* expression state. (c) *Esrrb* expression states mapped onto endpoints of lineage trees reconstructed by MEMOIR suggest that these states are stable for multiple generations. Two example colonies are shown. Scale bars, 20 μ m. (d) Frequency of occurrence in the same state (E^- , top; E^+ , bottom) of pairs of sisters, first cousins, and second cousins from MEMOIR

reconstructions of the 30 colonies with highest reconstruction confidence scores among the 85 colonies in which *Esrrb* was measured (blue, red) and from the actual lineages of the same colonies (gray). Transition rates inferred from MEMOIR are shown at right. (e–g) Cellular event recording (schematic). (e) gRNA1 (orange) is constitutively expressed for lineage reconstruction, while the orthogonal gRNA2 (purple) and gRNA3 (green) are expressed in response to specific signals and target independent scratchpads sets (SP). (f) Schematic showing recording of possible signaling histories (purple and green shading indicates periods when signals 1 and 2, respectively, are present). (g) Reconstruction of simulated event histories in a 6-generation tree. The signals recorded along two branches (yellow) are shown (bottom panels), including the actual simulated signals (thick lines), examples of individual reconstructed signals (dashed lines), and the average reconstructed signals (solid lines; mean \pm s.d., $n = 500$ trees) (Methods).

METHODS

MEMOIR component construction. The scratchpad transposon was constructed from a 10 repeat array (20X PP7 stem loops) derived from plasmid pCR4-24XPP7SL¹² and ligated directionally using BamHI and BglII sites into a modified form of the *piggybac* (PB) vector PB510B (SBI) lacking the 3' insulator and including a multiple cloning site (MCS). The CMV promoter was then removed using NheI and SpeI and replaced by a PGK promoter with Gibson assembly. A gBlock (IDT) containing the AvrII and XhoI restriction sites, priming sequences, and the sv40 polyA was then introduced 3' of the PP7 array by Gibson assembly using the EagI site in the backbone. Unique barcodes were then inserted into the transposon in the region 3' of the scratchpad array either by Gibson assembly or directed ligation using AvrII and XhoI. A total of 28 unique barcode sequences (Supplementary Table 1, GenScript Biotech Corp.) derived from *Saccharomyces cerevisiae* were used to generate the barcoded scratchpads. Scratchpad transposons were found to produce transcripts with half-lives of ~2 hours (Extended Data Fig. 1e–g).

The Cas9 construct was made using hSpCas9 from pX330⁷. First, the FKBP degenon (DD) was PCR-amplified from pBMN FKBP(DD)-YFP¹⁴ and introduced with Gibson assembly into pX330 restricted with AgeI, 5' of the open reading frame of hSpCas9, to create pX330-DD-hSpCas9. DD-hSpCas9 was amplified from this plasmid by PCR and introduced into another plasmid, 5' of a PGK promoter using Gibson assembly. After sequence verification, the PGK-DD-hSpCas9 construct was excised using restriction enzymes (AvrII and SacII), blunted with T4 polymerase, and ligated into a modified form of the *piggybac* vector PB510B (SBI) lacking the CMV promoter and including a MCS. A non-transposon version of Cas9 was also created using hSpCas9 amplified from pX330 and introduced with Gibson assembly at the 3' end of a CMV promoter containing two Tet operator sites into a standard plasmid backbone.

The Wnt pathway-responsive gRNA expression transposon was created using a LEF-1 response element¹⁵. The enhancer and promoter combination exhibited low basal activity, large dynamic range, and responsiveness to the GSK3 inhibitor Chir 99021 and the Wnt3a ligand. This Wnt sensor was cloned upstream of a nuclear localization signal (NLS)-tagged mTurquoise2, which served as a reporter of guide expression, that contained an embedded gRNA. The gRNA was flanked by self-cleaving ribozymes to excise it from the mRNA^{31,32}, and was purchased as a gBlock (IDT) and inserted using Gibson assembly between the end of the mTurquoise coding sequence and a SV40 polyA. This construct was contained in a modified form of the *piggybac* vector PB510B.

The Cre-activated gRNA expression transposon was created using the U6 TATA-lox promoter design³³, as illustrated (Extended Data Fig. 3a). The promoter, shRNA against mTurquoise, and gRNA regions were purchased as a gBlock or oligos (IDT) and inserted into a modified form of the *piggybac* vector PB510B containing PGK-H2B-mTurquoise.

Cell line engineering and culture conditions. To create MEM-01 we co-transfected the E14 mouse embryonic stem cell line (ATCC cat no. CRL-1821) with expression plasmids for hSpCas9 and the Tet repressor and then selected on neomycin. A single Cas9 positive clone was then used for co-transfection of 28 PB transposon barcoded scratchpads and a PB transposon PGK-palmitoylated-mTurquoise2/HygroR to facilitate segmentation of cell membranes and selection on hygromycin. Subsequent scratchpad-containing clones were inspected for overall scratchpad expression by smFISH. Scratchpad clones were also assessed for Cas9 expression, which was found to be very low and heterogeneous in most clones, with no expression in many cells (e.g., 6 ± 21 transcripts/cell). A scratchpad clone with good scratchpad expression was then simultaneously transfected with the DD-hSpCas9 PB transposon (to improve Cas9 expression (26 ± 17 transcripts/cell)) and the Wnt-activated gRNA expression PB transposon. Cells were selected on blasticidin. Single clones were assessed for activation potential on the basis of mTurquoise2 expression in response to CHIR99021 (Stemgent) or Wnt3a (1324-WN-002 R&D systems), and enhanced Cas9 expression was measured by smFISH. Among these clones was MEM-01, which demonstrated good gRNA activation in response to Wnt3a and increased Cas9 activity in the presence of the stabilizing agent, Shield1 (Clontech) (Extended Data Fig. 2c). MEM-01 resembled the parental E14 line in terms of cell morphology, cycle times, and expression of pluripotency markers including *Esrrb*, *Nanog*, and *SSEA-1*. Stably selected MEMOIR lines containing a Cre-activated gRNA were similarly engineered (Extended Data Fig. 3a–d).

The transfections described above were carried out using Fugene HD (Promega) at a mass (μ g) DNA/volume (μ L) Fugene ratio of 1:3 and following the manufacturer's instructions. For transfection of the PB components a total DNA mass of 1 μ g was used at a ratio of 6:1, PB transposons to PB transposase PB200PA-1 (SBI). For selection with antibiotics, transfected cells were lifted with Accutase (ThermoFisher) after transfection media was removed and plated on 100mm plates

(Nunc). 24 hours later growth media was replaced with selection media. Single colonies were lifted from selection plates as they matured.

During standard cell culturing, cells were maintained at 37 °C and 5% CO₂ in GMEM (Sigma), 15%, ES qualified fetal bovine serum (FBS) (Gibco/ThermoFisher), PSG (2 mM L-glutamine, 100 units/ml penicillin, 100 ug/ml streptomycin) (ThermoFisher), 1 mM sodium pyruvate (ThermoFisher), 1000 units/ml Leukemia Inhibitory Factor (LIF, Millipore), 1X Minimum Essential Medium Non-Essential Amino Acids (MEM NEAA, ThermoFisher) and 50 μ M – 100 μ M β -Mercaptoethanol (Gibco/ThermoFisher). Cells were maintained on polystyrene (Falcon) coated with 0.1% gelatin (Sigma).

Quantitative PCR. For detection of genomic barcode copy number, genomic DNA was prepared from cells using the Blood and Tissue kit (Qiagen). DNA was quantified on a NanoDrop 8000 spectrophotometer (ThermoScientific). Reactions were assembled as above with ~1,000–5,000 haploid genome copies, based on 3 picogram/haploid genome approximation. For gene expression analysis, total RNA was prepared using the RNeasy mini kit (Qiagen). One microgram of total RNA was used with the iScript cDNA synthesis kit (BioRad) following the manufacturer's instructions. For qPCR a 1:20 dilution of the cDNA was used in each reaction. All reactions were performed with IQ SYBR Green Supermix (BioRad). Reaction cycling was carried out on a BioRad CFX96 thermocycler. Both genomic DNA and cDNA samples were compared against *SdhA* copy number or expression level, respectively. Analyses included at least three biological repeats with each reaction run in triplicate, unless otherwise noted. Primers sets for all barcodes and normalizers were obtained from IDT, and the efficiencies of all primer pairs were tested.

Time-lapse movies and cell culture for imaging. Tissue culture grade glass bottom 24 well plates (MatTek) were treated with laminin-511 (20 μ g/ml) (Biolamina) for 4 hours at 37 °C and plated with cells at ~2500 cells/cm². Cells were exposed to Wnt3a (50–100 ng/ml) and Shield1 (50–100 nM) at the time of plating. After ~16 hours, cells were selected for time-lapse imaging based on system activation, assessed by visible mTurquoise2 signal, and then imaged in an incubated microscope environment every 14 min over 20–40 hours before being immediately fixed. Samples were fixed with 4% formaldehyde in PBS for 5 min. Samples cultured for smFISH imaging, but without time-lapse movie tracking, were prepared similarly (typically with a higher plated cell density) and activated for different lengths of time, as stated.

Single molecule fluorescence in situ hybridization (smFISH). Hybridization and imaging were carried out as previously described²³ with the following exceptions: scratchpad transcripts were targeted with 48 DNA oligo 20mer probes and barcode regions were targeted with 18 20mer probes (Table S2). Probes were coupled to one of three dyes (Alexa 555, 594, or 647 (ThermoFisher)) and used at ~130 nM concentration per probe set. Post-hybridization, cells were washed in 20% formamide in 2X SSC at 30 °C for 30 min, rinsed in 2X SSC, containing DAPI, at room temperature, and imaged in 2X SSC. For seqFISH, after imaging each round of hybridization, 2X SSC was replaced with wash buffer for ~5 min at room temperature and then replaced with the next probe set in hybridization buffer for overnight incubation. Most barcode signal from the previous hybridization were no longer visible during imaging of the following hyb (due to photobleaching and probe loss facilitated by the small number of barcode probes (18) used per barcode); any remaining visible transcripts were computationally subtracted during analysis. Incubation, washing, and imaging proceeded as above for up to 9 rounds of hybridization.

For analysis of smFISH images, semi-automated cell segmentation and dot detection were performed using custom Matlab software. Raw images were processed by a Laplacian-of-Gaussian filter and then thresholded to select dots. Colocalization between dots in the scratchpad image and barcode image was detected if both dots were above the threshold and within a few pixels of each other. To generate the histogram of intensities for the collapsed and uncollapsed scratchpads in Fig. 2b, we integrated the fluorescence intensities in the regions of the scratchpad smFISH image that corresponded to individual barcode dots. For the collapse rate experiment in Fig. 2c and Extended Data Fig. 3c, we measured the aggregate smFISH scratchpad colocalization levels for four highly expressed barcodes in cells that had been induced for different lengths of time. For activating conditions shown in Fig. 2b,c, only data from cells that were actually activated based on mTurquoise2 expression were included.

Lineage reconstruction of experimental data. Cell-to-cell barcode distance scores were determined for each pair of cells based on the similarity of the two cells' colocalization fractions for each barcode and weighted by the barcode's transcript number (as a measure of confidence in the observation). See Supplementary Information for details.

Lineage trees were reconstructed from the cell-to-cell barcode distance matrices using a modified version of a standard agglomerative hierarchical clustering algorithm³⁴. Reconstructions were constrained to binary trees such that cells were

paired into sisters before first cousin pairs were assigned. Pairing proceeded by successively grouping pairs of cells or cell clusters with the minimum barcode distance. At each step, if the two most optimal (i.e., minimum distance) pairings were close in distance, the algorithm optimized for the lowest combined distance of the current and next minimum distances. The distance between two clusters was computed using the standard UPGMA algorithm¹⁹ by averaging the cell-to-cell barcode distance between all possible pairs of cells across the two clusters.

Bootstrap to identify robust reconstructions. For each colony, the barcoded scratchpad data were resampled by bootstrap and a corresponding lineage tree was reconstructed ($n = 1000$ resampled reconstructions per colony). Based on the frequency at which the original cousin clades occurred in the resampled reconstructed trees, a robustness score was assigned to each colony. Colonies whose clade reconstructions were less sensitive to resampling showed significantly improved overall reconstruction accuracy. Subsets of colonies with more reliable reconstructions could thus be selected without prior knowledge of their accuracy by selecting colonies with higher robustness scores, e.g., scores in the top 20–40% of the data.

Alternative metrics for identifying colonies with robust lineage information were also tested. These metrics similarly enriched for subsets of data with improved reconstruction accuracy, further supporting the observation that some colonies showed clear lineage information while others did not acquire well-defined collapse patterns likely due to limited, excessive, or ambiguous collapse events.

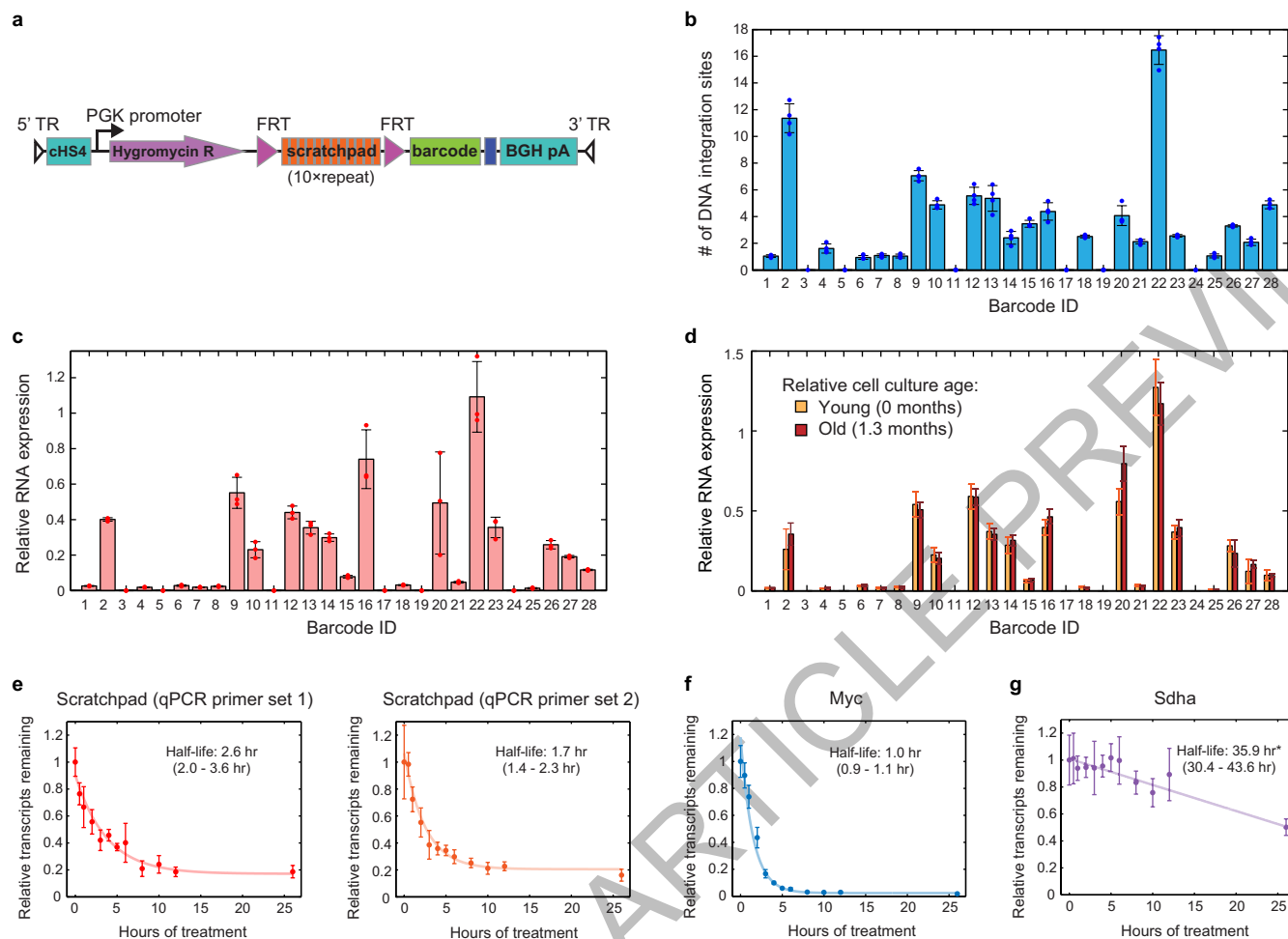
Lineage reconstruction simulations. To simulate MEMOIR for three generation binary trees, we started with one cell with a fixed number of idealized scratchpads. At each division, the daughter cells inherited the same scratchpad profile as their parent and independently collapsed each uncollapsed site with a fixed probability, defined as the collapse rate. After three generations, the scratchpad profiles of the 8 resulting cells were used to reconstruct their lineage tree using either a modified neighbor joining algorithm³⁴, or the Camin-Sokal maximum parsimony algorithm³⁵ that exhaustively scored all 315 possible tree reconstructions. Both forward simulations and the reconstruction algorithms were implemented in Matlab. For the heatmap and the cumulative distribution functions shown in Fig. 3g–i, the fraction of correct relationships was computed as the fraction of all distinct pairwise relationships in the actual tree that were correctly identified in the reconstructed tree. If multiple reconstructions were equally valid (same parsimony score), the fraction of correct relationships was averaged over all of them. Reconstruction accuracy was tested over a wide range of collapse rates (Fig. 3h) or for the approximate collapse rate observed in our experiments, 0.1 per site per generation (Fig. 3g,i). The empirical collapse rate, 0.1, was estimated from the observed colocalization fraction of the barcodes, ~ 0.67 , in 108 MEM-01 colonies induced for ~ 48 hrs (same colonies as in Fig. 3). In Extended Data Fig. 8a, trees of a higher number of generations were reconstructed from the final collapse pattern using a modified neighbor joining algorithm³⁴ in which allowed reconstructions were restricted to full binary trees. Fraction of correct relationships was again computed as the fraction of all distinct pairwise relationships in the actual tree that were correctly identified in the reconstructed tree averaged over at least 1000 trees.

Event recording simulations. *Simulation of signal recording.* To demonstrate event recording, we simulated the same forward tree generation algorithm as in the MEMOIR lineage reconstruction simulations (Fig. 3h and Methods), for trees of 6 generations, assuming 50 idealized scratchpads and a collapse rate of 0.1 per scratchpad per generation. The simulated cells also contained two additional sets of recording scratchpads of 50 sites each (Fig. 4e). We assume these scratchpads collapse through independent events occurring at rates proportional to the magnitude of their respective input signals. The minimum and maximum collapse rates at low and high signal were set to 0 and 0.2 per scratchpad per generation, respectively. The magnitude of the input signals varied over time and from branch to branch as shown in Fig. 4g, resulting in different collapse rates for each of the two recording scratchpad sets over time and along different lineages.

Reconstruction of simulated signal dynamics. We first reconstructed the lineage tree using only the lineage tracking scratchpad sites. This reconstruction used a neighbor joining algorithm, as in Fig. 3h³⁴. We then reconstructed the history of the collapse events of the recording scratchpads on the reconstructed lineage tree. For this procedure, we used a Camin-Sokal maximum parsimony algorithm³⁵. Briefly, the algorithm proceeds from the leaves of the tree to the root. At each generation, it infers the collapse state of the parental node, based on the known collapse states of the two daughters, while minimizing the number of new collapse events occurring between the parent and the daughters. For binary scratchpads this corresponds to computing the intersection between the collapse patterns of the two daughters. This procedure is then repeated for the parent and its sister until reaching the root. At the end of this procedure, one obtains a maximum parsimony assignment of scratchpad states to each node in the tree. Based on these assignments, we calculated the number of scratchpad collapse events in recording scratchpads that occurred along each branch. Finally, this reconstructed collapse rate provides an estimate of the underlying signal intensity along each lineage (e.g., actual and reconstructed signals shown for two lineages of interest in Fig. 4g).

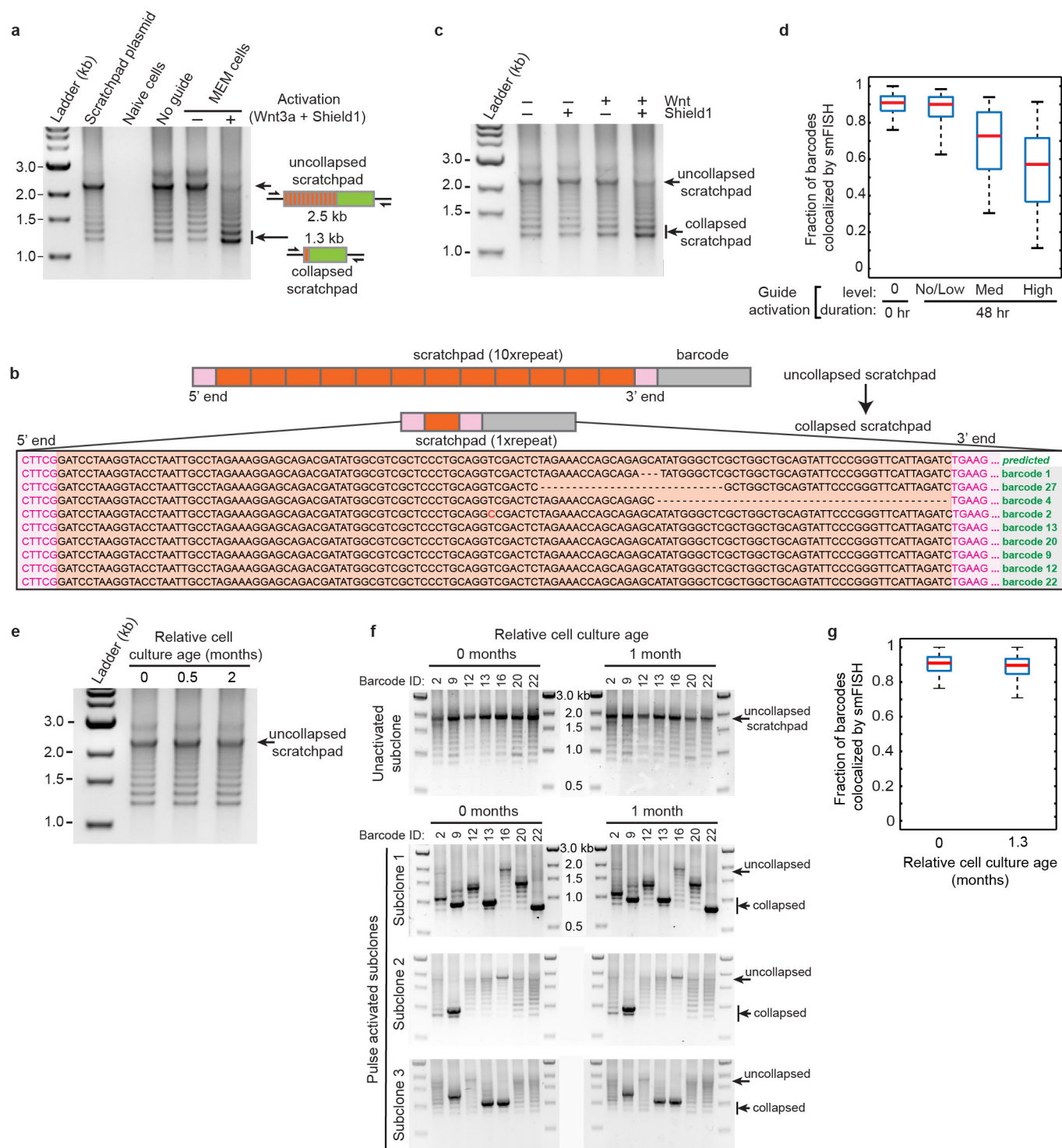
Data availability. Data that are not included in the paper are available upon reasonable request to the authors.

- Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **56**, 343–349 (2014).
- Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**, 698–710 (2014).
- Ventura, A. et al. Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10380–10385 (2004).
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- Camin, J. H. & Sokal, R. R. A Method for deducing branching sequences in phylogeny. *Evolution* **19**, 311–326 (1965).
- Chung, J. H., Whiteley, M. & Felsenfeld, G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74**, 505–514 (1993).
- Sharova, L. V. et al. Database for mRNA half-life of 19,977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* **16**, 45–58 (2009).
- Friedel, C. C., Dölken, L., Ruzsics, Z., Koszinowski, U. H. & Zimmer, R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res.* **37**, e115 (2009).
- Clark, M. B. et al. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
- Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).
- Percharde, M. et al. Nco3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming. *Genes Dev.* **26**, 2286–2298 (2012).
- Uranishi, K., Akagi, T., Sun, C., Koide, H. & Yokota, T. Dax1 associates with Esrrb and regulates its function in embryonic stem cells. *Mol. Cell. Biol.* **33**, 2056–2066 (2013).
- Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
- Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).
- Sankoff, D. Minimal mutation trees of sequences. *SIAM J. Appl. Math.* **28**, 35–42 (1975).
- Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
- Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).



Extended Data Figure 1 | MEM-01 consistently expresses short-lived transcripts from multiple integrated barcoded scratchpads. (a) The barcoded scratchpad transposon is composed of the following elements (left to right): the *piggyBac* 5' terminal repeat (triangle), the chicken HS4 insulator³⁶, a PGK promoter driving expression of the hygromycin resistance coding sequence, a 5' FRT site, the PP7 scratchpad array consisting of 10 repeats, a 3' FRT site, a barcode sequence (Supplementary Table 1), a priming region for sequencing and PCR, the BGH polyA, and the *piggyBac* 3' terminal repeat (triangle). (b) Unique genomic integrations for the MEM-01 cell line were detected by qPCR. Bars show mean \pm s.d. of 4 biological repeats with individual data points marked. (c) The relative RNA expression levels of barcode integrations were quantified by RT-qPCR. Bars show mean \pm s.d. of 3 biological repeats with individual data points marked. (d) Scratchpad expression profiles remain constant over 1.3 months of passaging. Low and high passage cultures of MEM-01 cells (light and dark bars, respectively) were assayed for RNA expression levels by RT-qPCR. The unchanged expression levels indicate that most barcoded scratchpads express at a consistent level and are not routinely silenced over time. Bars show values from single biological samples with error bars calculated by combining in quadrature the technical

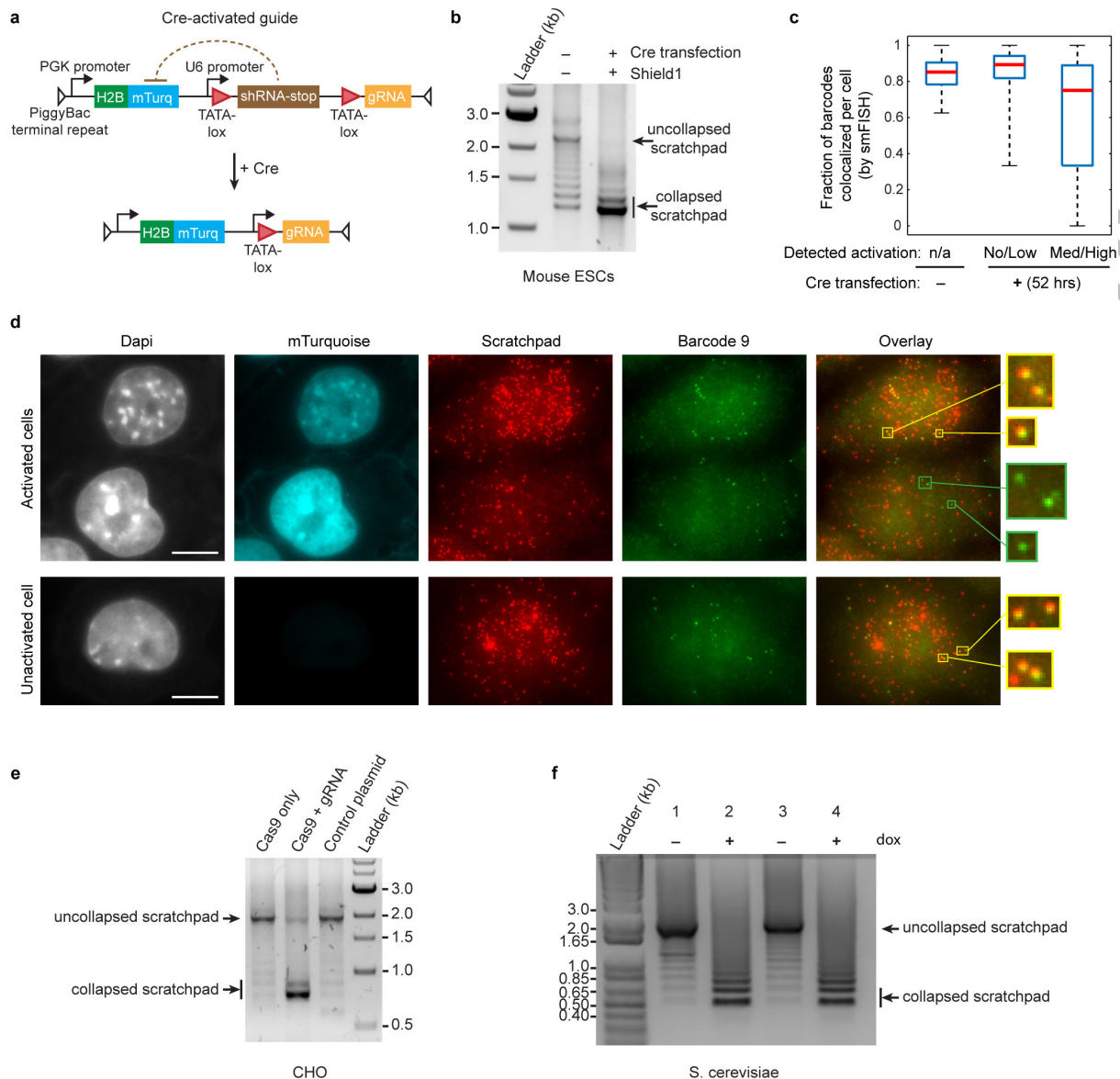
replicate variation in barcode and normalizer Cq values. (e-g) RNA half-lives assessed by RT-qPCR analysis of transcript levels after blocking transcription with actinomycin D (10 μ g/ml). (e) Barcoded scratchpad transcripts were assayed with two different sets of qPCR primers (left and right panels). These data indicate a half-life of \sim 2 hr. (f,g) Myc and Sdha are known to have short and long mRNA half-lives, respectively, and were assessed as controls, for comparison³⁷⁻³⁹. (f) Myc half-life of 1 hr was shorter than the other measured lifetimes, while (g) Sdha was longer lived. For Sdha, the measured half-life value (*) is expected to overestimate the true value, as Sdha levels were determined relative to those of the similarly long-lived gene *Atp5e*, whose transcript levels were also decaying over the time course. A previous estimate of Sdha half-life in mESCs was \sim 8-13 hr³⁷. All sample transcript levels were assessed relative to those of *Atp5e*³⁷⁻³⁹. Transcript abundances were normalized to 1 at time zero. Decay curves were fit assuming one-phase exponential decay using weighted nonlinear least squares regression (e,f) or assuming a linear approximation to exponential decay (g). Half-lives were determined based on the best fit decay constants and a range reported based on the 95% confidence interval (parenthesis)). Data represent two biological replicates with multiple technical replicates; error bars show standard deviations.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Barcoded scratchpads collapse to truncated products in activated cells and are stable in full length and collapsed forms. (a) Agarose gel electrophoresis of PCR amplified scratchpads reveals scratchpad collapse after gRNA induction. Full-length scratchpads were amplified from plasmid DNA (lane 1), as well as from cells without gRNA constructs (lane 3), or with uninduced gRNAs (lane 4). By contrast, cells expressing gRNA showed shorter products (lane 5). Cells with no scratchpads are also shown as a negative control (lane 2). Bands corresponding to the full length scratchpad and the collapsed scratchpad are indicated (arrows). Note: The laddering effect seen in all lanes and gels is due in part to PCR amplification artifacts with the repetitive arrays. For gel source data, see Supplementary Figure 1. (b) The lowest molecular weight band from scratchpad collapse, as shown in lane 5 in a, was extracted and subcloned into a vector. Nine of the colonies were sequenced. They aligned to a single repeat unit with 5' and 3' flanking regions, suggesting complete collapse of the repeats due to Cas9 activity. 6 of the 9 sequencing reads resulted in collapse to a perfect single repeat (with a point mutation in the barcode 2 sequence), and the remaining 3 sequencing reads had additional small deletions in the scratchpad. (c) Scratchpad collapse requires induction of both Cas9 and gRNA. Gel shows scratchpad states for MEM-01 cells treated with no ligand, with Shield1 (to stabilize Cas9 protein), with Wnt3a (to induce gRNA expression), and with both Wnt3a (100 ng/mL) and Shield1 (100 nM), all after 48h. (d) Scratchpad collapse increased with increasing gRNA activation, as assessed using smFISH to detect scratchpad colocalization

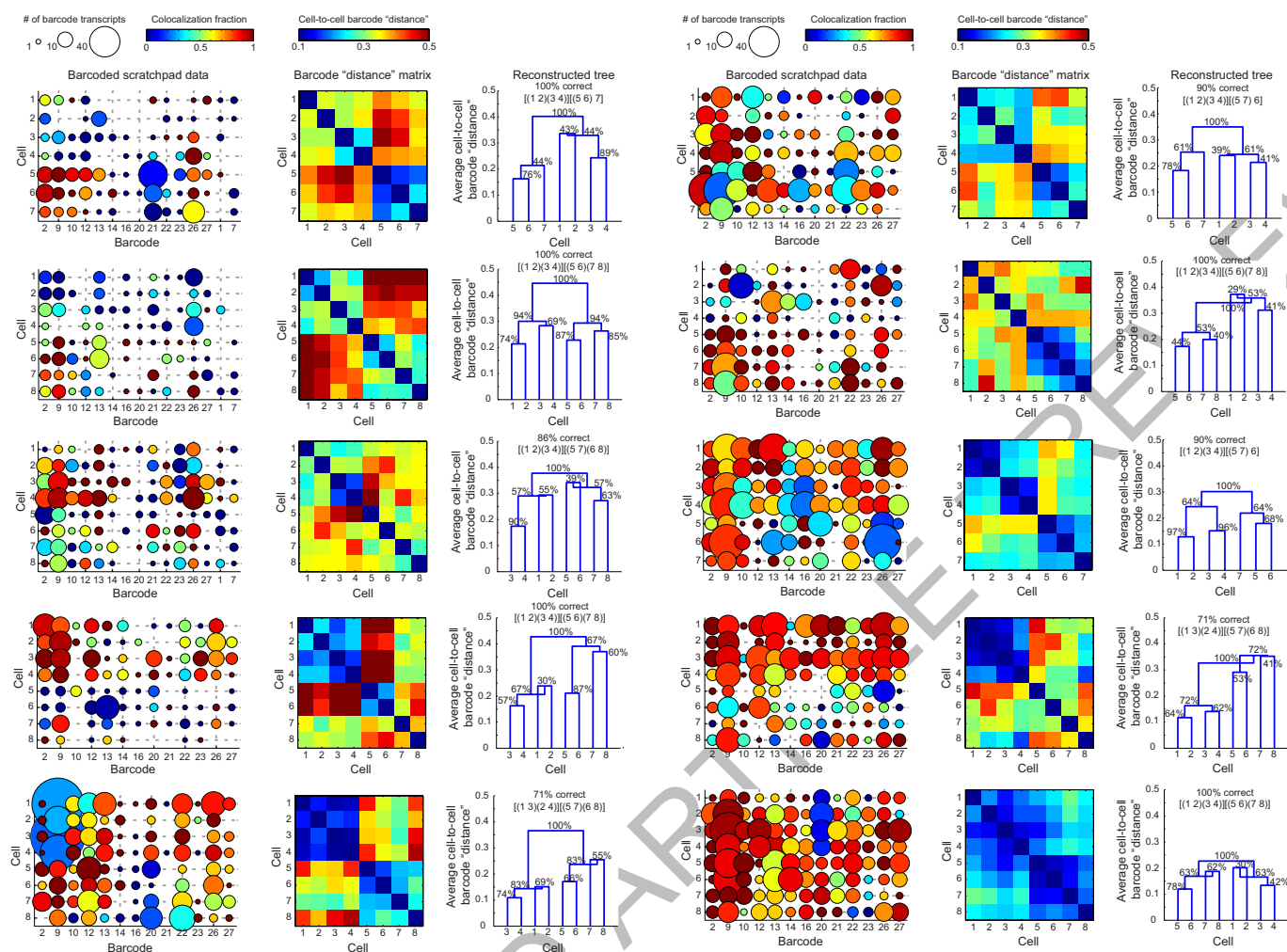
with 4 highly expressed barcodes. Cells were analyzed either without gRNA activation or 48 hours after gRNA activation by addition of Wnt3a and Shield1 (same concentrations as in c). gRNA expression was measured by the intensity of co-expressed nuclear mTurquoise signal. Boxplots show median (red bar), 1st and 3rd quartiles (box), and extrema of distributions; $n = 1826, 1081, 345, 191$ cells, left to right. Related to Figure 2c. (e-g) Scratchpad states remain stable over extended periods. (e) Unactivated MEM-01 cells maintained uncollapsed scratchpads over timescales of months. (f) To check the stability of individual barcoded scratchpad variants over time, multiple subclones of MEM-01 were isolated after no activation (control; top panels) and after a pulse of activation for 24 hours (Wnt3a 100 ng/mL, Shield-1 100 nM; bottom panels). Subclones were assessed for the states of different barcoded scratchpad types after initial isolation (0 month relative age, left) and after one month of maintenance (right). The apparent collapse states (from uncollapsed to fully collapsed) of the barcoded scratchpad types were (1) distinct in different subclones, and (2) remained stable over a month, indicating that scratchpad states are stable over these timescales. (g) Barcoded scratchpads are also stable over long periods as assessed by smFISH readout. The fraction per cell of barcode transcripts (from 4 distinct barcode types) that colocalized with scratchpad signal was essentially unchanged between an unactivated low passage cell culture and one maintained for over a month. The imperfect colocalization fraction is largely the result of errors in smFISH detection and not gradual scratchpad collapse. Boxplots as in (d); $n = 1826, 983$ cells, left to right.



Extended Data Figure 3 | Scratchpad collapse works with an alternative gRNA, and in multiple cell types.

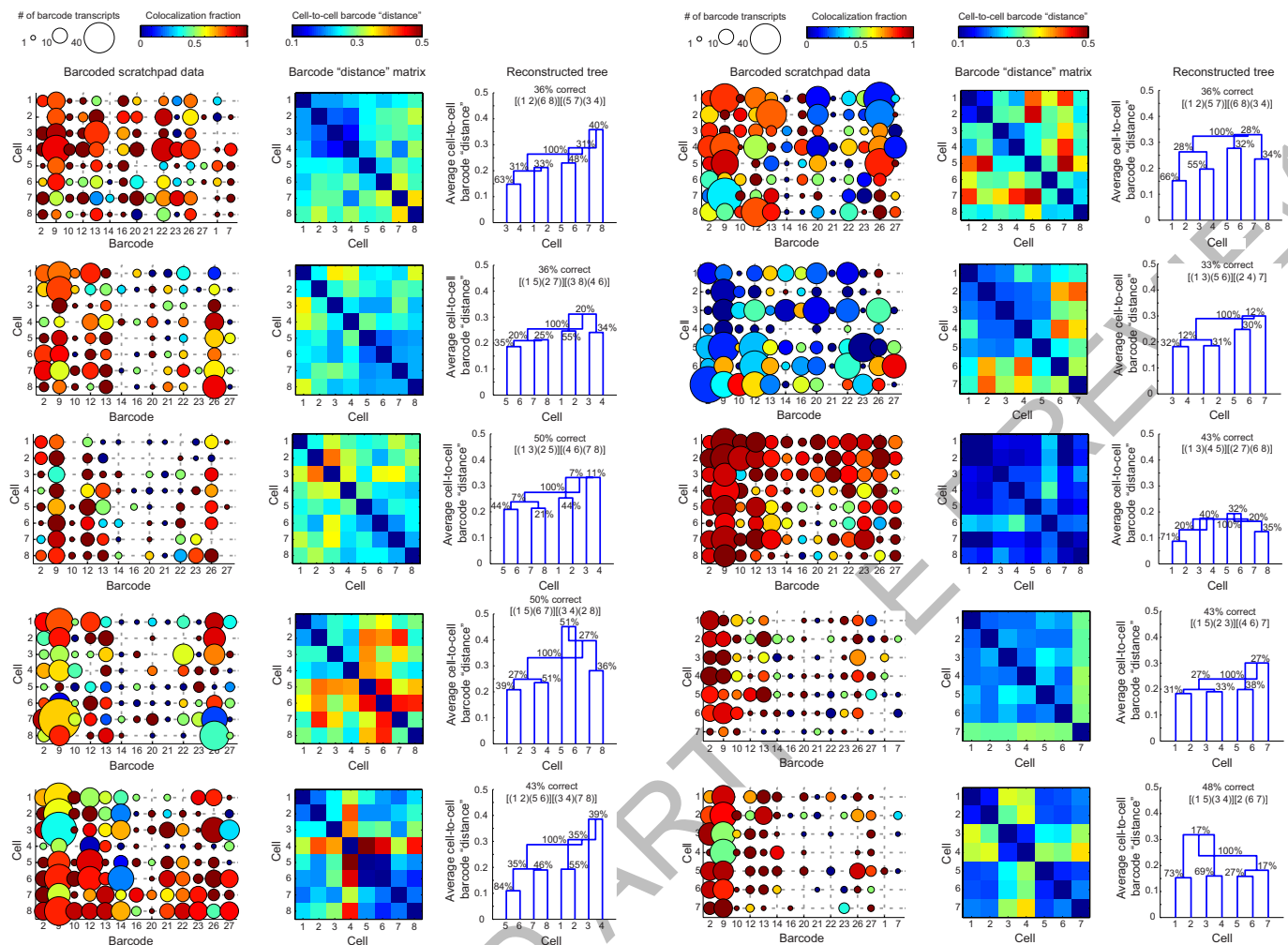
(a-d) A Cre recombinase-activated gRNA is effective at inducing collapse events. (a) Schematic of Cre-activated gRNA system. The construct contains a constitutive PGK promoter driving expression of a Histone 2B (H2B)-mTurquoise fusion protein (the H2B provides nuclear localization). This is followed by a U6 TATA-lox promoter³³ driving expression of an shRNA against mTurquoise, followed in turn by a polyT (T6) transcriptional stop, and then a gRNA directed against scratchpad regions. Prior to Cre expression, expression of the shRNA keeps mTurquoise levels low (brown dashed line) and prevents expression of the gRNA. After activation of Cre, the shRNA-stop cassette is removed, allowing mTurquoise and gRNA expression. Thus, mTurquoise provides a visual marker of gRNA expression. This type of gRNA architecture would allow MEMOIR activation in specific tissues expressing Cre. (b) PCR analysis shows that Cre can induce scratchpad collapse. Gel shows genomic DNA from a clonal cell line harboring the construct in a. Scratchpads appear uncollapsed in untransfected cells (left lane), but show significant collapse after transfection with mRNA encoding Cre protein (right lane, ~52 hr after transfection). Note: The laddering effect seen in all lanes and gels is due in part to PCR amplification artifacts with the repetitive arrays. (c) smFISH analysis reveals Cre-activated scratchpad collapse. Quantification of barcode-scratchpad colocalization fractions as measured by smFISH. Cre transfection reduced scratchpad and barcode colocalization levels in cells that expressed mTurquoise, indicating Cre activity (right). Cells that were mTurquoise negative or low and untransfected cells retained high colocalization levels (middle and left).

Colocalization levels per cell were assessed based on the colocalization of four expressed barcodes with scratchpad transcripts. Boxplots show median (red bar), 1st and 3rd quartiles (box), and extrema of distributions; n = 995, 643, 649 cells, left to right. (d) Example smFISH images of scratchpad and barcode colocalization detected in single cells containing the Cre-activated gRNA. Some activated cells (top panels, mTurquoise expression "on") show loss of colocalized signal for a specific barcode (top panels, lower cell). Unactivated cells, as assessed by low mTurquoise expression, typically show no loss of colocalization (bottom panels). Scale bars, 10 μ m. (e-f) Scratchpads in CHO-K1 cells and yeast also undergo Cas9/gRNA-dependent collapse. (e) Cas9 and gRNA expressing plasmids were transiently transfected into Chinese Hamster Ovary (CHO-K1) cells containing stably integrated scratchpads. Gel analysis reveals Cas9 and gRNA dependent scratchpad collapse (middle lane), while transfection with Cas9 expression plasmid alone or control plasmids resulted in no collapse (left and right lanes, respectively). (f) Scratchpad collapse was tested in a yeast strain with doxycycline inducible Cas9 and gRNA and integrated scratchpads. Before inducing Cas9-gRNA expression (lane 1 and 3, (-) dox), the scratchpads were intact. After Cas9-gRNA induction with 2 μ g/ml doxycycline for 11 hours, scratchpads appeared collapsed (lane 2 and 4, (+) dox). Left 2 lanes (lanes 1 and 2) and right 2 lanes (lanes 3 and 4) correspond to 2 biological replicates. Note that the scratchpads in CHO-K1 and yeast cells have a similar scratchpad PP7 array to that used elsewhere but different flanking sequences, so their absolute PCR product lengths differ. For gel source data, see Supplementary Figure 1.



Extended Data Figure 4 | Examples of lineage reconstruction for 10 colonies. Data for 10 colonies that reconstructed with >70% of pairwise relationships correctly identified are shown here. The bubble-chart shows the number of barcode transcripts detected (bubble size) and the uncollapsed fraction (color scale). Matrix of cell-to-cell barcode "distance" (dissimilarity) scores were computed from the data. Low (blue) values indicate more similar barcoded scratchpad collapse patterns. Note that sisters and cousins tend to have lower distance scores than second cousins, creating a block diagonal pattern in the distance matrix. Lineage trees

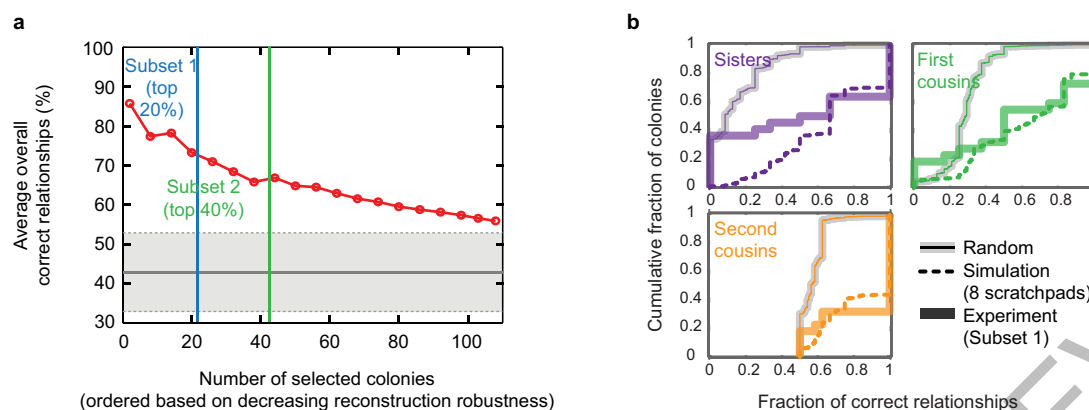
were reconstructed based on the distance matrix using an agglomerative hierarchical clustering algorithm (see Methods). Cluster distances from the reconstruction algorithm are shown as branch heights in the reconstructed linkage trees. Percentages on the linkage trees represent frequencies of clade occurrence from a barcode resampling bootstrap. The percentage of correct relationships identified by the depicted lineage reconstruction is shown (xx% correct) and the actual tree is reported as [(x y)(x y)][(x y)(x y)], where sister pairs are denoted as (x y) and cousins are grouped in brackets ([...]).



Extended Data Figure 5 | Analysis of reconstruction failure modes.

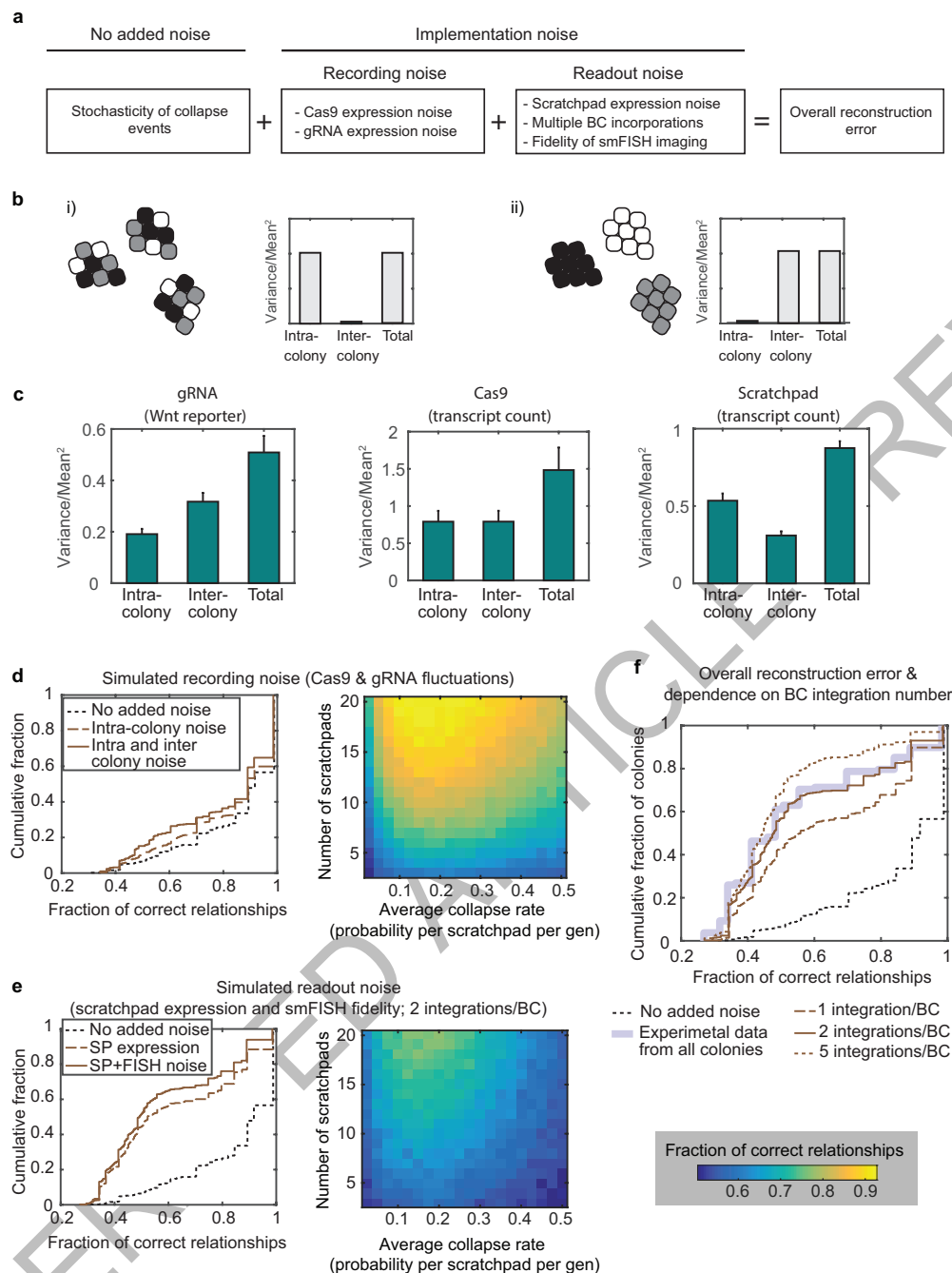
These 10 colonies showed reconstruction accuracies similar to those of random data. Bubble charts, distance matrices and linkage trees are shown as in Extended Data Figure 4. Note the relative lack of block diagonal structures in the distance matrices, which typically reflect evidence of close sister or cousin relationships and less similar second cousins in better reconstructed colonies. Poor reconstructions result from insufficiently informative or inconsistent collapse patterns. These can occur in several ways: First, colonies may have too many collapsed scratchpads (e.g., row 2, col 2), leading to degeneracy, and eliminating differences between clades. Second, and more often, colonies have too few collapsed scratchpads (e.g., row 3, col 2) to reconstruct the full tree accurately. Third, colonies can provide inconsistent or incomplete lineage information such that the data do not point to one consistent lineage hypothesis (e.g., row 5, col 1). Inconsistent information can arise from convergent collapse events in which the same scratchpad randomly collapses in separate branches of the lineage—such noise is inherent to this method of lineage tracking

but can be significantly reduced by increasing the number of barcoded scratchpads. Additionally, variability in scratchpad expression, resulting from stochastic expression of individual barcoded scratchpads as well as apparent inconsistencies due to expression of multiple incorporations of the same barcoded scratchpad can generate conflicting information. Despite these issues, colonies can in many cases provide information about some lineage relationships. For example, for the colony in row 5, col 1, all the sister pairs are correctly identified, but they are not definitively placed in the lineage tree due to conflicting readouts at the cousin level (e.g., collapse events in barcodes 9 and 14). Similarly, for the colony in row 5, col 2, cells 3 and 4 are readily identified as sisters because of a common collapse event in barcode 9. But, there is little additional information, such as a collapse event from the 2-cell-stage, which would allow the cousins to be correctly identified. These and other sources of noise impacting colony reconstruction are analyzed in more detail in Extended Data Fig. 7 and Supplementary Information, and can be addressed in future implementations of MEMOIR.



Extended Data Figure 6 | Bootstrap reconstruction score enriches for colonies that exhibit more accurate lineage reconstruction. (a) A bootstrap procedure (Methods) was used to determine the robustness of clade reconstruction to resampling of barcode data for each colony. The frequency of lineage reconstruction at the first cousin clade level was then used to rank all 108 colonies. Colonies with higher reconstruction robustness were enriched for more accurate lineage reconstructions, although no information about accuracy was used to identify these colonies. The top 20% of colonies based on bootstrap score were termed subset 1 (left of blue line; $N = 22$). This group correctly identified an average of 72% of relationships. The top 40% of colonies were termed subset 2 (left of green line; $N = 43$) and correctly identified 67% of relationships. Gray region indicates the range of correct relationships expected from random guessing of trees (mean \pm s.d. indicated by line

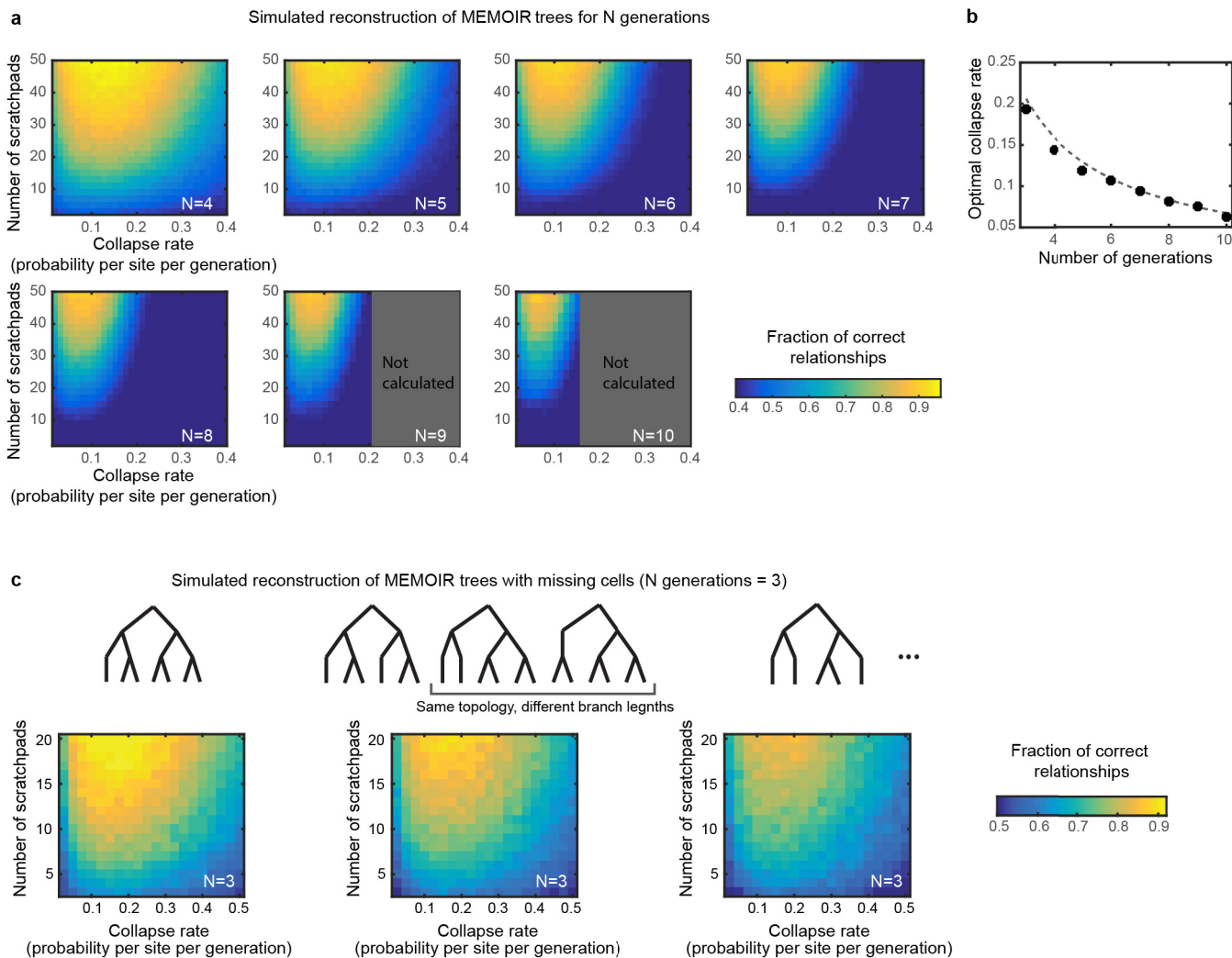
and shading). The bootstrap metric effectively filters out colonies that have insufficient or inconsistent scratchpad collapse information and thus do not robustly generate the same reconstruction. Noise sources that affect the data include convergent scratchpad collapse, imperfect collapse rates that may not result in collapse events every generation, and variable scratchpad expression that limits readout signal or introduces ambiguities due to expression from multiple incorporations of the same barcode type (see Extended Data Fig. 7 and Supplementary Information). (b) Cumulative distributions show the fraction of pairwise sister, first cousin, and second cousin relationships correctly identified in each colony. Reconstruction accuracies of all these types of lineage relationships are similar to predictions based on the simulated model with 8 scratchpads (no noise included). This shows that reconstruction is accurate across all levels of relationships. Related to Fig. 3g.



Extended Data Figure 7 | Comprehensive error analysis identifies scratchpad expression variability as the key source of noise in MEMOIR experiments.

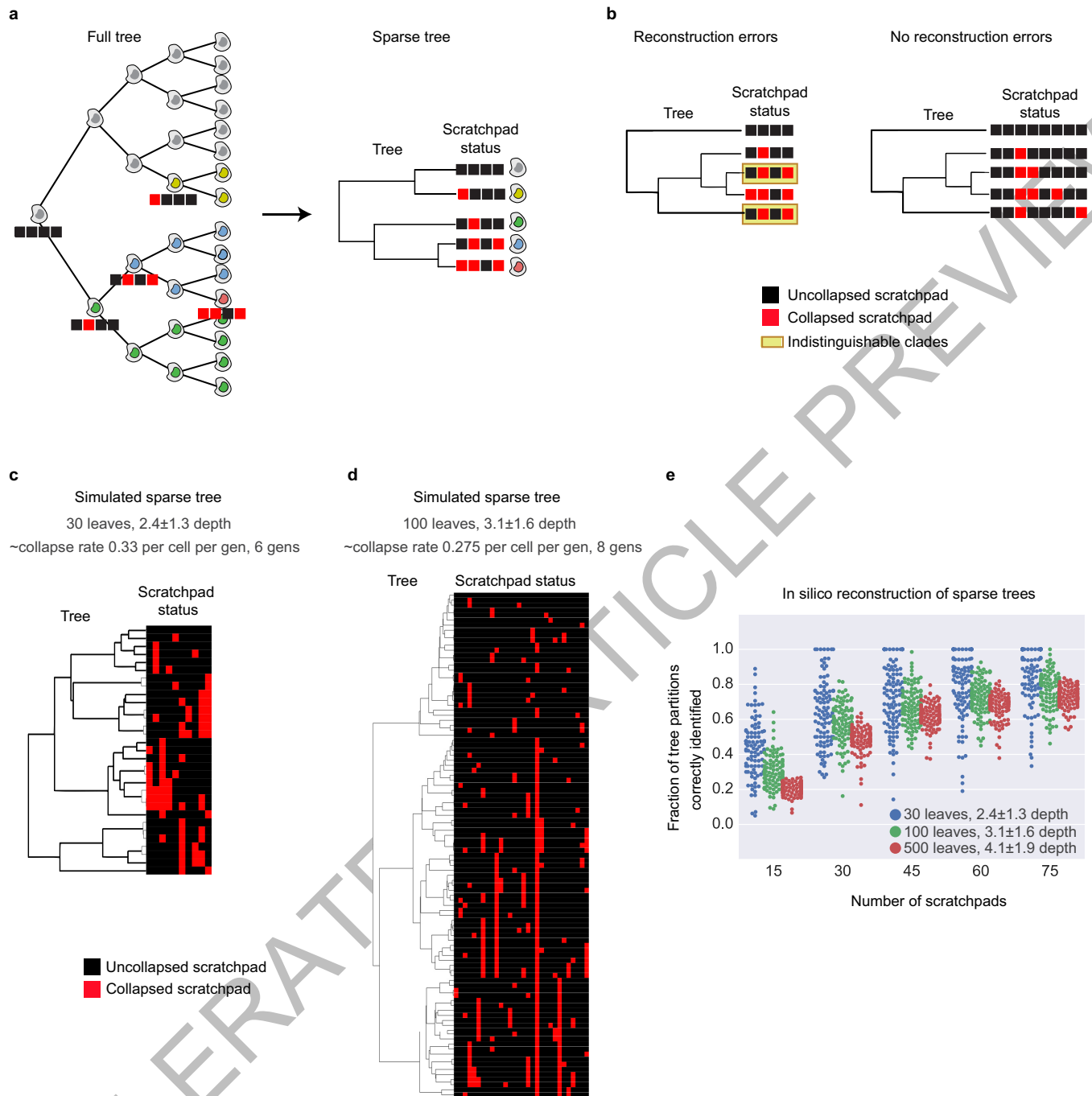
(a) Overall reconstruction errors result from three types of noise: the inherent stochastic nature of recording lineage information with stochastic scratchpad collapse events, recording noise (due to fluctuations in the expression levels of Cas9 and gRNA), and readout noise (due to fluctuations in the expression levels of the barcoded scratchpads, variable expression from multiple integrations of the same barcoded scratchpad species (BC), and the fidelity of smFISH imaging readout). (b) Cell-cell variability can be decomposed into intra-colony and inter-colony components, as shown schematically. For each hypothetical colony, the relative amounts of each type of variability are plotted (also schematic). (c) Plots show experimentally measured intra- and inter-colony noise from gRNA activity (from the fluorescent signal of the Wnt reporter, left), Cas9 expression (from the transcript counts by smFISH, middle), and scratchpad expression (from transcript counts by smFISH, right). These plots represent data from individual cells of all 108 MEM-01 colonies (see Supplementary Information for details). (d) Recording noise results in a small decrease in reconstruction accuracy. The plot on the left shows the

cumulative distribution of reconstruction accuracies of 500 simulated colonies comprised of trees of 3 generations, with an average scratchpad collapse rate of 0.1, and 13 scratchpads. The heatmap on the right shows the average reconstruction accuracy for 500 simulated colonies for a range of average collapse rates and number of scratchpads. (e) Fluctuations in scratchpad (SP) expression levels significantly reduce reconstruction accuracy. Simulation results are plotted as in d, but with the addition of readout noise, rather than recording noise, to the idealized simulations. The readout noise is added as two separate components: scratchpad expression level fluctuations, which significantly increase error, and noise due to smFISH imaging fidelity, which contributes minimally to reconstruction error. The curves are for 2 integration sites per barcode. (f) Cumulative distribution of reconstruction accuracy of 500 simulated colonies with all three components of noise included for different numbers of integration sites per barcode. The thick blue line is the experimental distribution obtained from the 108 MEM-01 colonies. The simulated distribution is consistent with the experimentally observed distribution, especially for two effective integrations per barcode. (No fitting parameters were used.)



Extended Data Figure 8 | Performance analysis on deeper trees and trees with missing cells. (a) Simulations of reconstruction accuracy of full binary trees for varying numbers of unique barcoded scratchpads, varying collapse rates, and varying numbers of generations (N). The color of the heat maps corresponds to the fraction of all pairwise lineage relationships correctly identified in the reconstructed tree, averaged over many simulated trees (Figure 3h in the main text, also see Methods). Even at greater depth (e.g., $N = 10$), trees can be reconstructed accurately with ~ 50 scratchpads. (b) The collapse rate that maximizes reconstruction accuracy depends on the number of generations to be tracked, but is only weakly dependent on the number of scratchpads. This is because maximal lineage information is recorded when each scratchpad has a 0.5 probability of having collapsed by the final time point, regardless of the total number of scratchpads. The plot shows the optimal collapse rate as a function of tree depth, as determined from the simulations (dots) as well as the theoretical expectation of a 0.5 cumulative collapse probability per scratchpad (dashed line). The theory curve contains no fitting parameters. (c) Simulations of reconstruction accuracy for binary trees of three generations as a function of the number of scratchpads and the scratchpad

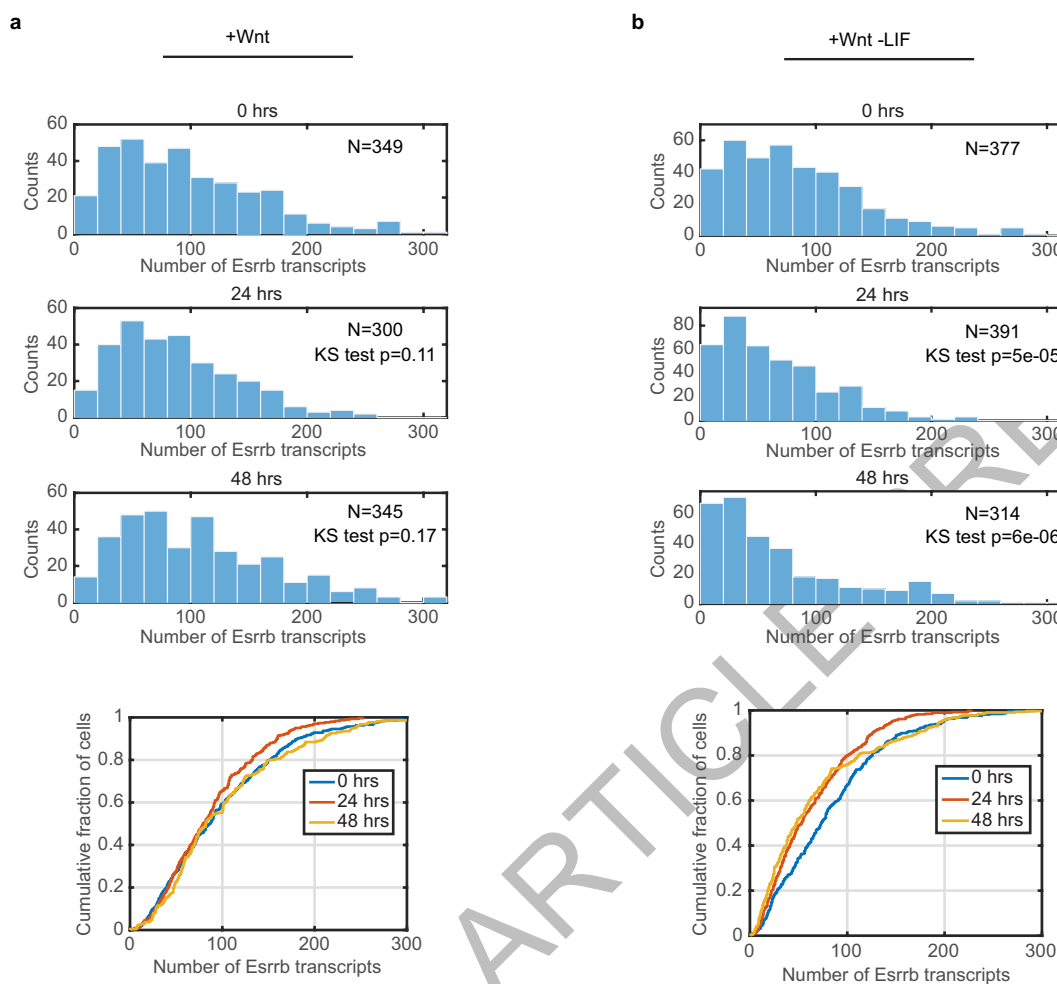
collapse rate for trees with one (left), two (middle), or three (right) randomly chosen endpoint cells missing. (Compare with reconstruction accuracy for trees with no missing cells in Figure 3h.) The schematic above each panel shows the topology and branch lengths of trees with the given number of missing cells. A modified neighbor joining algorithm³⁴ was used to exhaustively score all 315 possible reconstructions. To distinguish between reconstructions where tree topology is the same but the branch lengths are different (two such trees are shown bracketed in the schematic of the middle panel), we modified the reconstruction algorithm to estimate the branch lengths connecting a pair of cells based on the hamming distance of their barcoded scratchpad collapse patterns (see Supplementary Information). For example, two cells whose collapse patterns differ significantly would be estimated to have a longer lineage distance between them than would cells with more similar patterns. In general, trees with missing leaves can be reconstructed with accuracy similar to full binary tree (Fig. 3h). As the number of missing cells increases, the reconstruction accuracy decreases because there are fewer cells in the tree to provide lineage information.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Simulations show that MEMOIR can operate at low collapse rates to reconstruct sparse trees. We simulated MEMOIR in the sparse recording regime, in which collapse events for any given lineage occur, on average, once every few generations. Trees were generated using simulations and reconstructed using a maximum parsimony approach (see Supplementary Information). (Experimentally, sparse tree regimes in which collapse events occur infrequently could be achieved with low Cas9 and/or gRNA expression levels or rare expression events (e.g., by using weak promoters, occasionally-activated promoters, protein degradation domains), or with decreased Cas9-mediated affinity for target scratchpads (e.g., by decreasing the complementarity between the gRNA and target). (a) Cartoon of sparse collapse events on a full binary tree. Each collapse changes the state of each scratchpad (arrays of red or black boxes, shown only at nodes where new collapse events occur). At the final generation, there are five populations of cells with distinct collapse patterns, each shown in a different color. In the sparse representation of the tree (right) each collapse event corresponds to a new branch, and the five leaves correspond to the five subpopulation of cells with distinct collapse patterns. (b) Possible source of reconstruction errors. Unrelated clades can converge independently to the same collapse pattern and thus become indistinguishable, resulting in reconstruction errors (tree on the left), but the probability of such coincidences decreases with increasing number of scratchpads (all clades are distinguishable

for the tree on the right). (c) A simulated sparse tree with 30 leaves and an average depth of 2.4 ± 1.3 . The depth of the tree is defined as the cumulative number of collapse events experienced by each leaf averaged over all the leaves of the tree. The statistics of this tree shape is approximately equivalent to a sparse tree generated by a collapse rate of 0.33 per cell per generation on a full tree of 6 generations. The heatmap shows the status of the scratchpad sites for all the leaves. Each column corresponds to a particular barcoded scratchpad, and each row to a leaf. (d) Same as in (c), but for a simulated sparse tree with 100 leaves and a depth of 3.1 ± 1.6 ; approximately equivalent to a collapse rate of 0.275 per cell per generation on a full tree of 8 generations. (e) The fraction of correctly identified tree partitions (defined using the Robinson-Foulds metric⁴⁰) is shown as a function of the number of scratchpads, and normalized by its value in the limit of an infinite number of distinct scratchpads (where a unique collapse pattern is generated for every collapse event). Sparse trees of three different sizes (i.e., number of leaves and depth) were generated. Each dot corresponds to one simulated tree. Tree size was held constant as the number of scratchpads was increased, requiring a fixed collapse rate per cell but a collapse rate per scratchpad that scaled inversely with scratchpad number. Trees with fewer leaves and lower depth required fewer scratchpads for accurate reconstruction. But, even larger trees could recover close to the maximal lineage information using only a modest number of scratchpads.



Extended Data Figure 10 | The *Esrrb* expression level distribution is stationary. (a) Distribution of the number of *Esrrb* transcripts in individual cells in populations of MEM-01 ES cells activated by the addition of Wnt3a and Shield1 (same conditions as the colonies analyzed in Figures 3 and 4) for different amounts of time (0 hrs, 24 hrs, and 48 hrs, respectively from top to bottom). The distribution of *Esrrb* transcript counts does not change significantly over 48 hours of Wnt3a exposure as quantified by the p-value of the Kolmogorov-Smirnov (KS) test. The KS test was performed for the observed distributions at 24 and 48 hours

with respect to the reference distribution at 0 hours. The cumulative distribution functions (bottom) similarly show that the fraction of cells in the low (or high) *Esrrb* expression state does not change significantly over 48 hours of Wnt activation. A stationary *Esrrb* distribution implies that transitions between the low and high *Esrrb* expression states must be reversible. (b) LIF removal changes the *Esrrb* distribution. Same as in panel a but with LIF removed from the media at $t=0$. The distributions show a significant change during the 48 hours, with the fraction of cells in the low *Esrrb* expression state increasing over time, as expected^{41,42}.