RESEARCH ARTICLE SUMMARY

DEVELOPMENTAL BIOLOGY

Lineage tracing on transcriptional landscapes links state to fate during differentiation

Caleb Weinreb*, Alejo Rodriguez-Fraticelli*, Fernando D. Camargo†, Allon M. Klein†‡

INTRODUCTION: During tissue turnover, stem and progenitor cells differentiate to produce mature cell types. To understand and ultimately control differentiation, it is important to establish how initial differences between cells influence their ultimate choice of cell fate. This challenge is exemplified in hematopoiesis, the ongoing process of blood regeneration in bone marrow, in which multipotent progenitors give rise to red cells of the blood, as well as myeloid and lymphoid immune cell types.

In hematopoiesis, progenitor cell states have been canonically defined by their expression of several antigens. However, as in several other tissues, recent transcriptome analysis by single-cell RNA sequencing (scSeq) showed that the canonically defined intermediate cell types are not uniform, but rather contain cells in a variety of gene expression states. scSeq also showed that the states of hematopoietic progenitors form a continuum, differing from classic depictions of a discrete stepwise hierarchy.

ON OUR WEBSITE

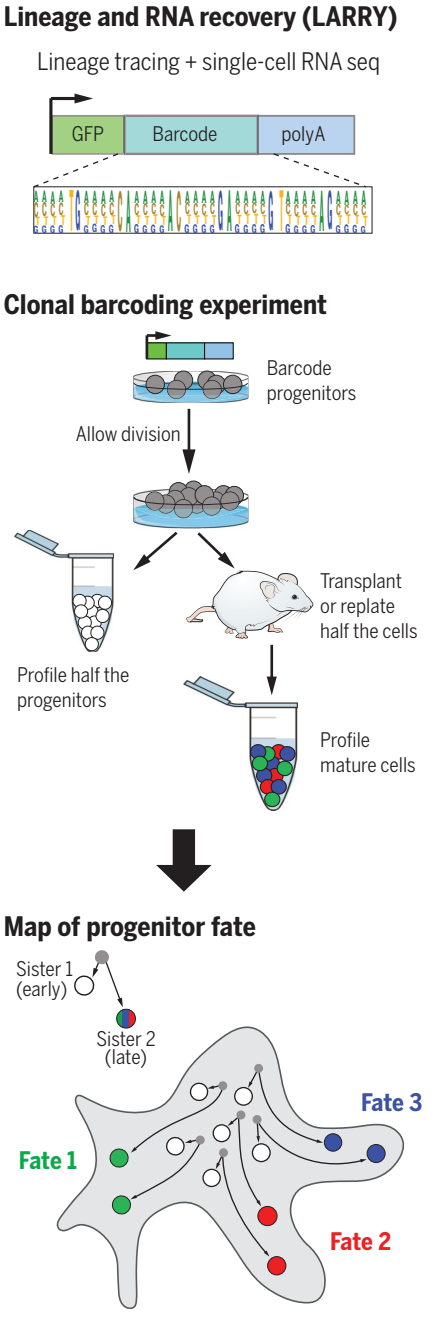
Read the full article at http://dx.doi.org/10.1126/science.aaw3381

RATIONALE: In this study, we set out to establish how variation in transcriptional state biases future cell fate and whether scSeq is sufficient to completely distinguish cells with distinct fate biases. Directly linking whole-transcriptome descriptions of cells to their future fate is challenging because cells are destroyed during scSeq measurement. We therefore developed a tool we call LARRY (lineage and RNA recovery) that clonally tags cells with DNA barcodes that can be read using scSeq. Using LARRY, we aimed to reconstruct the genome-wide transcriptional trajectories of cells as they differentiate.

RESULTS: We linked transcriptional progenitor states with their clonal fates by barcoding heterogeneous cells, allowing cell division, and then sampling cells for scSeq immediately or at later time points after differentiation in culture or in transplanted mice. We profiled >300,000 cells in total, comprising 10,968 clones that gave information on lineage relationships at single time points and 2632 clones spanning multiple time points in culture or in mice. We confirmed that clonal trajectories over time approximated the trajectories of single cells and were thus able to identify states of primed fate potential on the continuous transcriptional landscape. From this analysis, we identified genes correlating with fate, established a lineage hierarchy for hematopoiesis in culture and after transplantation, and revealed two routes of monocyte differentiation that give rise to distinct gene expression programs in mature cells. The data made it possible to test state-of-the-art algorithms of scSeq analysis, and we found that fate choice occurs earlier than predicted algorithmically but that computationally predicted pseudotime orderings faithfully describe clonal dynamics.

We investigated whether there are stable cellular properties that have a cell-autonomous influence on fate choice yet are not detected by scSeq. By analyzing clones split between wells or transplanted into separate mice, we found that the variance in cell fate choice attributable to cell-autonomous fate bias was greater than what could be explained by initial transcriptional state. Less formally, sister cells tended to be far more similar in their fate choice than pairs of cells with similar transcriptomes. These results suggest that current scSeq measurements cannot fully separate progenitor cells with distinct fate bias. The missing signature of future fate choice might be detectable in the RNA that is not sampled during scSeq. Alternatively, other stable cellular properties such as chromatin state could encode the missing information.

CONCLUSION: By integrating transcriptome and lineage measurements, we established a map of clonal fate on a continuous transcriptional landscape. The map revealed transcriptional correlates of fate among putatively multipotent cells, convergent differentiation trajectories, and fate boundaries that could be not be predicted using current trajectory inference methods. However, the map is far from complete because scSeq cannot separate cells with distinct fate bias. Our results argue for looking beyond scSeq to define cellular maps of stem and progenitor cells and offer an approach for linking cell state and fate in other tissues. ■



Lineage and transcriptome measurements allow fate mapping on continuous cell state landscapes. A tool we named LARRY labels cell clones with an scSeq-compatible barcode. By barcoding cells, letting them divide, and then sampling them immediately or after differentiation, it is possible to link the initial states of cells with their differentiation outcomes and produce a map of cell fate bias on a continuous transcriptional landscape.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: Allon_Klein@hms.harvard.edu
Cite this article as C. Weinreb *et al.*, *Science* **367**, eaaw3381 (2020); DOI: 10.1126/science.aaw3381.

RESEARCH ARTICLE

DEVELOPMENTAL BIOLOGY

Lineage tracing on transcriptional landscapes links state to fate during differentiation

Caleb Weinreb[1]*, Alejo Rodriguez-Fraticelli[2,3]*, Fernando D. Camargo[2,3]†, Allon M. Klein[1]†‡

A challenge in biology is to associate molecular differences among progenitor cells with their capacity to generate mature cell types. Here, we used expressed DNA barcodes to clonally trace transcriptomes over time and applied this to study fate determination in hematopoiesis. We identified states of primed fate potential and located them on a continuous transcriptional landscape. We identified two routes of monocyte differentiation that leave an imprint on mature cells. Analysis of sister cells also revealed cells to have intrinsic fate biases not detectable by single-cell RNA sequencing. Finally, we benchmarked computational methods of dynamic inference from single-cell snapshots, showing that fate choice occurs earlier than is detected by state-of the-art algorithms and that cells progress steadily through pseudotime with precise and consistent dynamics.

During differentiation, stem and progenitor cells progress through a hierarchy of fate decisions, refining their identity until reaching a functional end state. The gold standard for inferring the relationship between progenitors and their offspring is lineage tracing, in which a subset of progenitors is labeled, typically using genetic approaches that mark cells expressing defined marker genes, and their fate is profiled at a later time point (*1*). Lineage maps are key to understanding and controlling differentiation (*2*).

Recently, whole-genome approaches for profiling cells by single-cell RNA sequencing (scSeq) introduced a complementary approach to understanding developmental relationships. scSeq captures mature cell types along all stages of cell differentiation, revealing a "state map" in gene expression space. These state maps offer hypotheses for the hierarchy of cell states (*3*) and their gene expression dynamics over time (*4*–*7*). Unlike lineage tracing, scSeq can be performed without prior genetic manipulation and without being limited by the specificity of transgene expression within the progenitor cell pool (*2*).

Neither state nor lineage mapping alone, however, provides a complete view of the differentiation processes. Whereas scSeq offers a very high resolution of cell states, it cannot link the detailed states of progenitors to their ultimate fate because cells are destroyed in the process of measurement. scSeq data do not directly report the stages at which progenitor cells become committed to one or more fates or how many distinct paths might lead cells to the same end states. In addition, the high-dimensional nature of scSeq allows more than one approach to constructing cell-state trajectories from the same data (*4*). There is a need for approaches that link the detailed whole-genome state of cells to their long-term dynamic behavior.

Here, we integrate measurements of cell lineage with scSeq using the mouse hematopoietic system as a model of fate choice. In adults, hematopoietic stem and progenitor cells (HSPCs) reside in the bone marrow and maintain steady-state blood production. Cell culture and transplantation studies over several decades have led to the prevailing model of hematopoiesis as a branching hierarchy with defined fate-restricted intermediates (*8*). However, recent state maps from scSeq (*9*), as well as clonal studies using barcodes (*10*) and single-cell culture (*11*), suggest that the traditional intermediate cell types are internally heterogeneous in state and fate potential, with HSPCs lying along a continuum of states rather than a stepwise hierarchy. Reconciling these views requires tracking the dynamics of individual lineages on the continuous landscape of HSPC states defined by scSeq (*12*). Here, we explore an experimental design for capturing the state of a cell at the whole-transcriptome level and its clonal fate at a later time point simultaneously across thousands of cells in different states.

RESULTS

Simultaneous assay of clonal states and fates

Our strategy for simultaneously capturing transcriptional cell state and fate is to genetically barcode a heterogeneous progenitor population, allow cell division, sample some cells immediately for scSeq profiling, and then sample the remaining cells later (*13*). This approach provides data for three types of clonal relationships (Fig. 1A): (type 1) sister cells in the earliest time point that may be captured after one or two rounds of division; (type 2) clones observed at both early and later time points that allow comparing the state of an early cell with the fate outcomes of its sisters; and (type 3) differentiated cells sampled at later time points that will reveal clonal relationships between different fates. If recently divided sister cells (type 1) are transcriptionally similar, then pairs of clonally related cells sampled both early and late (type 2) should reveal how single-cell gene expression changes over time during differentiation. This approach can map the fate of cells from a continuous landscape of starting states and does not require isolation or labeling of specific prospective progenitor populations (*2*, *14*).

We modified a classical strategy for clonal labeling by lentiviral delivery of inherited DNA barcodes (*15*, *16*) to allow barcode detection using scSeq (*17*). The barcode consists of a random 28-mer in the 3′ untranslated region of an enhanced green fluorescent protein transgene (*eGFP*) under control of a ubiquitous EF1α promoter (Fig. 1B). Transcripts of eGFP are captured during scSeq, and the barcode is revealed through analysis of sequencing reads. We generated a library of ~0.5×10^6 barcodes, sufficient to label 5000 cells in an experiment with <1% barcode overlap between clones (see materials and methods, section 2.3, for an estimate of diversity). We refer to the barcoding construct as LARRY (lineage and RNA recovery).

We tested LARRY on mouse embryonic stem cells and primary HPCs. After profiling by scSeq, one or more barcodes could be robustly detected in 93% of GFP^+ cells (fig. S1, a to c). Specific barcode sequences overlapped rarely between replicate transduction experiments at a frequency expected by chance for the library size (0.3% of 5000 barcodes appeared more than once). Therefore, the approach provides an efficient method for simultaneously barcoding large numbers of cells for combined fate and state mapping.

To analyze HSPC fate potential, we applied LARRY to cells cultured in vitro and cells transplanted in vivo. For in vitro analysis, we isolated a broad class of oligopotent ($Lin^-Sca^-Kit^+$) and multipotent ($Lin^-Sca1^+Kit^+$ or LSK) progenitor cells (fig. S2, a and b) and plated them in media chosen to support broad, multilineage differentiation (see the materials and methods). After barcode transduction, cells were cultured for 2 days to allow lentiviral integration and subsequent division. During this time, the cells divided three times on average. We then sampled half the cells (defining the "early state") for scSeq. The other half were replated and then sampled after 2 days (30% of cells) and 4 days (remaining cells) (Fig. 1C). For transplantation, $Lin^-Sca^{(hi)}Kit^+$ cells, consisting of mostly short-term and long-term hematopoietic stem cells (ST-HSCs and LT-HSCs, respectively) (fig. S2, a and b), were barcoded and placed in

[1]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [2]Stem Cell Program, Boston Children's Hospital, Boston, MA 02115, USA. [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: allon_klein@hms.harvard.edu

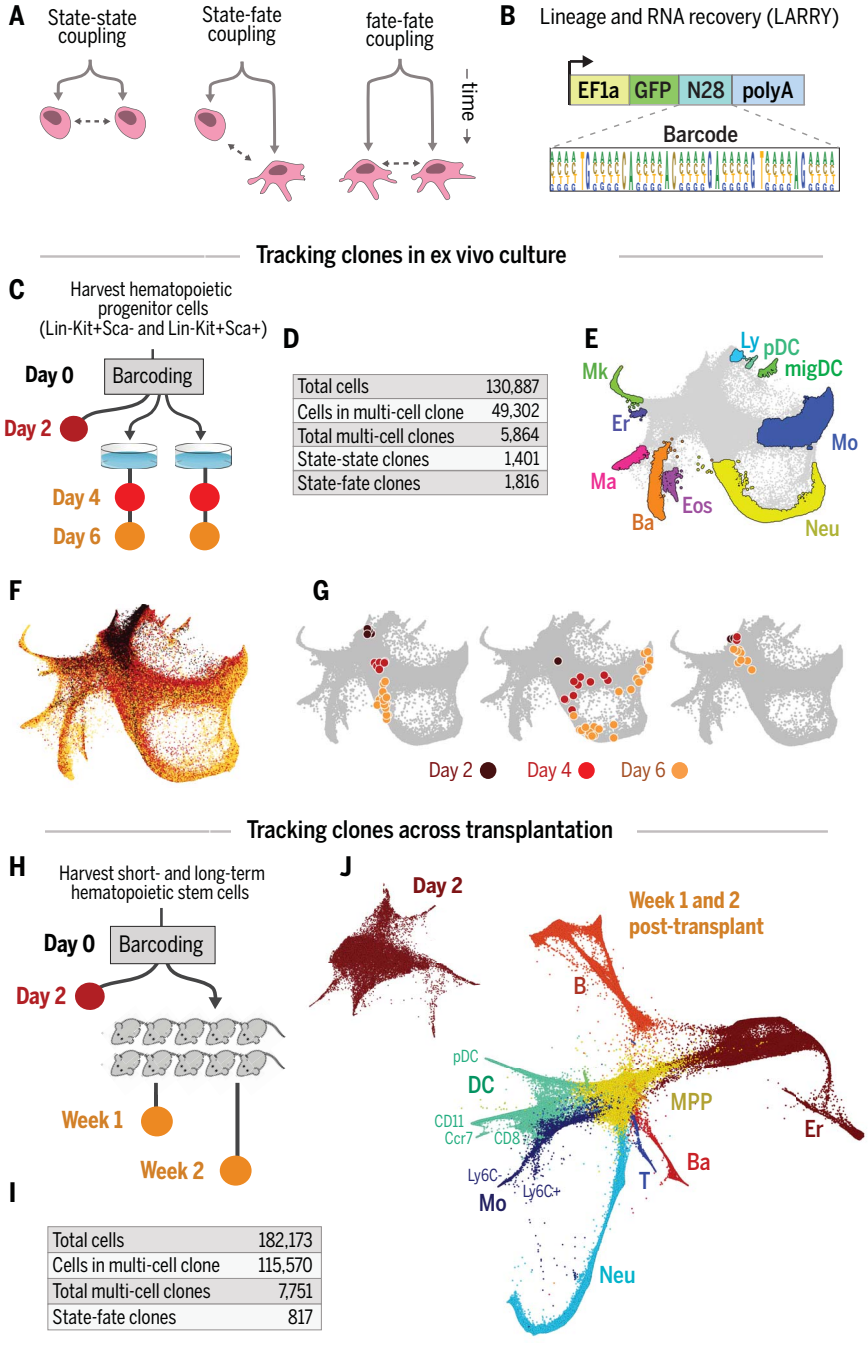


Fig. 1. Tracking clones over hematopoietic differentiation. (**A**) Experimental designs for tracking differentiation dynamics by analysis of sister cells. (**B**) The LARRY lentiviral construct delivers an expressed, heritable barcode that is detectable using scSeq. (**C** to **G**) Experiment tracking hematopoietic progenitor clones over time in primary culture. (C) Colored circles indicate samples collected for scSeq. (D) Numbers of cells and clones sampled. (E) Annotated SPRING plot of transcriptomes from all time points. (F) SPRING plot colored by the time point at which cells were profiled. (G) Examples of clonal dynamics on the single-cell landscape. Each plot shows a separate clone, with cells colored by time point and overlaid on the full dataset in gray. (**H** to **J**) Experiment tracking clones after transplantation into 10 mice. (H) Colored circles are as in (C). (I) Numbers of cells and clones sampled. (J) scSeq data before transplantation (top left) and after transplantation (bottom right) plotted as in (E). T, T cell; B, B cell; NK, natural killer cell.

culture. After 2 days, 40% were profiled by scSeq, with the remainder transplanted into 10 sublethally irradiated host mice (*10*) and recovered for scSeq 1 and 2 weeks later (Fig. 1H). We retrieved 130,887 scSeq transcriptomes from culture and 182,173 single cells after transplantation (see table S1 and materials and methods, section 3, for details of this analysis). In these two experiments, 38% and 63% of cells, respectively, belonged to a clone of two or more cells (5864 and 7751 clones), with 1816 and 817 clones in total spanning early and late time points (Fig. 1, D and I).

We visualized the cell transcriptomes using force-directed layouts [SPRING plots (*18*)]. In vitro, the cells defined a continuous state map spanning from multipotent progenitors (MPPs) to nine mature cell types that appeared in culture (Fig. 1, E and F): erythrocytes (Er), megakaryocytes (Mk), basophils (Ba), mast cells (Ma), eosinophils (Eos), neutrophils (Neu), monocytes (Mo), plasmocytoid dendritic cells (pDCs), $Ccr7^+$ migratory DCs (migDCs), and lymphoid precursors (Ly). On this landscape, clones exhibited a range of behaviors, including unilineage and multilineage differentiation and self-renewal of early progenitors (Fig. 1G). After transplantation, the cells again defined a continuous landscape spanning from MPPs through several stages of Neu maturation, as well as DCs, Mo, Er, B, T, and Ba cells. Many of these cell types were internally heterogeneous, with several types of DCs, including $CD11^+$, $CD8^+$, migDCs, and pDCs, as well as $Ly6C^+$ classical and $Ly6C^-$ nonclassical Mos (Fig. 1J and fig. S3). We did not detect Mks, possibly because they did not survive bone marrow harvest, flow sorting, and single-cell encapsulation intact. Therefore, with these experiments, we simultaneously captured

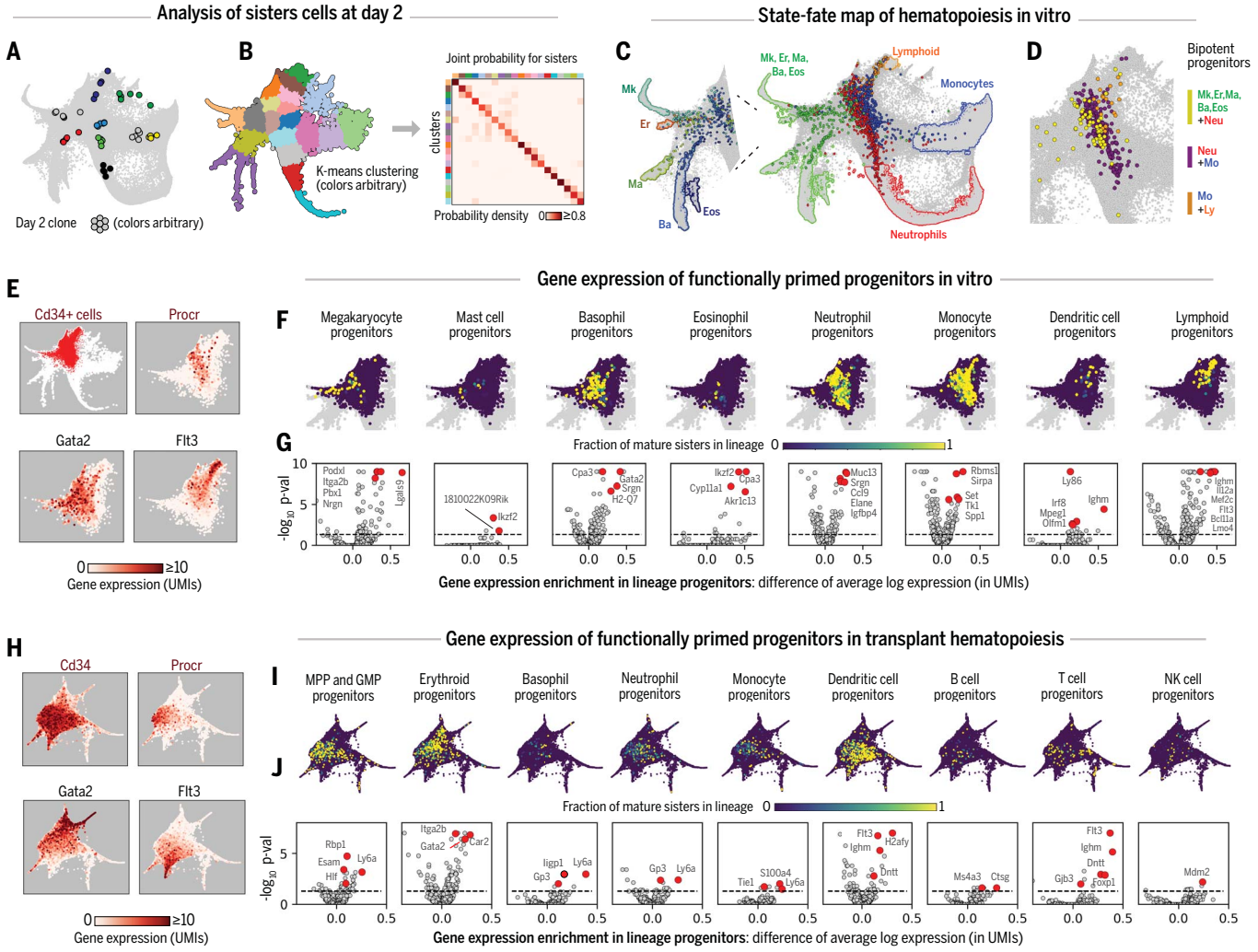


Fig. 2. Linking state to fate in early hematopoiesis. (**A** and **B**) Sister cells at day 2 are transcriptionally similar, as seen by example [(A); each color shows one clone] and by the probability of sister cells occupying the same or neighboring transcriptional clusters (B). (**C**) Day 2 cells (colored dots) are colored by the fate of their mature sisters observed at a later time in vitro. Outlined regions of the SPRING plot indicate the respective fates. (**D**) Location of progenitors (colored dots) with two fates among their sisters at later time points. (**E**) Gene expression domains of day 2 cells guide selection of early progenitors for further analysis. (**F**) Early progenitors colored by the fraction of sisters in each fate at days 4 to 6 in culture. (**G**) Volcano plots identifying genes enriched among early progenitors for each lineage. Labeled genes are shown in red. UMIs, unique molecular identifiers. (**H**) Heterogeneity among purified LSK cells after 2 days in culture before transplantation into mice. (**I** and **J**) Detection of early progenitor gene expression associated with future fates after transplantation, repeating analyses from (E) to (G). In (E), (F), (I), and (J), points with nonzero value are plotted on top.

single-cell state maps and their underlying clonal relationships.

Clonal dynamics identify early transcriptional fate boundaries

With LARRY, it is possible to estimate how a single cell changes over time by sampling a clone across multiple time points. However, the accuracy of this approximation depends critically on the similarity of sister cells at the earliest time point. We found that pairs of sisters profiled on day 2 localized in the SPRING graph had correlated gene expression (median $R = 0.846$) and that most (70%) fell in the same or nearest neighbor cluster (Fig. 2, A and B; fig. S4, a to d; materials and methods, section 5). A minority of cells, however, were more diverged, with 10% falling outside a four-cluster radius (compared with 80% for random cell pairs). We tested and ruled out that similar sister pairs were technical co-encapsulation artifacts (fig. S4e). These tests justified approximating single-cell trajectories by clonal trajectories, although with some loss in resolution of fate boundaries expected because of ~10% diverged sister pairs.

Beginning with the in vitro data, we recorded the clonal fates of each day 2 cell. Visualizing cells from unilineage clones revealed well-delineated domains of fate potential (Fig. 2C). Where the progenitors for different fates overlapped, we observed bipotent or oligopotent clones, indicating the location of fate commitment boundaries (Fig. 2D). The true number of multipotent clones is likely underestimated in our data because some clonal fates were likely missed due to undersampling (fig. S5) and cell commitment before division would result in only one observed fate. Consistent with recent scSeq studies (*19*), progenitors with different fate potentials did not partition into discrete cell states, but instead formed a structured continuum. Further, bipotent domains formed extended fate boundaries, indicating that differentiation progression can occur independently of fate commitment over some time. Both of these observations differ from the classical model of hematopoiesis represented

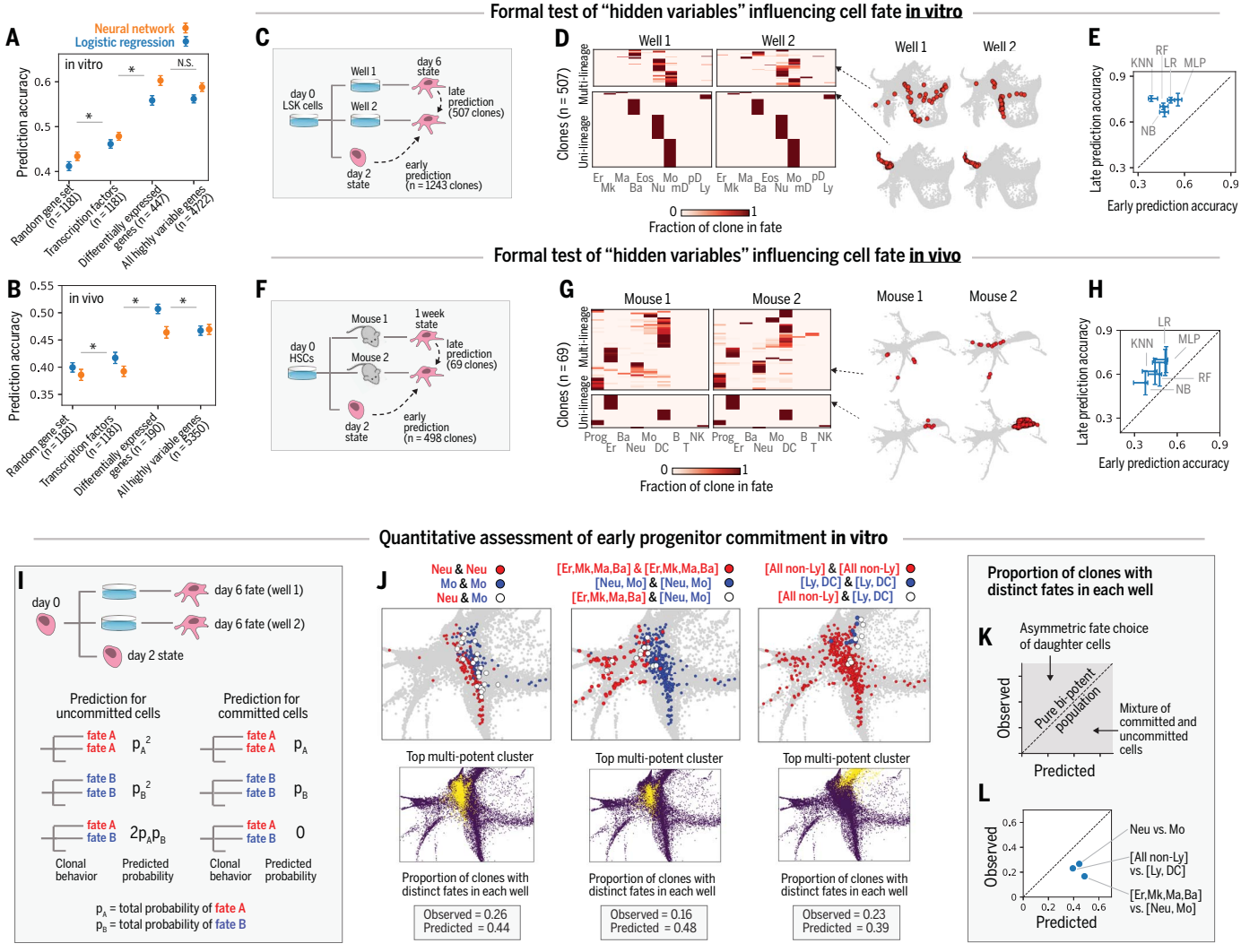


Fig. 3. Stochasticity and hidden variables from scSeq data. (**A** and **B**) Machine learning partially predicts clonal fate from the transcriptional state of early progenitors in vitro and in vivo. Accuracy is the fraction of correct assignments. Asterisk (*) indicates statistical significance ($p < 10^{-4}$). N.S., not significant. Error bars indicate standard deviation. (**C** and **F**) Split-well and mouse experiments testing for heritable properties that influence fate choice but are not detectable by scSeq. Hidden heritable properties are implicated if cell fate outcomes are better predicted by the late (day 6 in vitro, 1 week in vivo) state of an isolated sister cell compared with the early (day 2) state of a sister. (**D** and **G**) Clonal fate distributions for sister cells split into different wells or different mice and profiled on day 6. Each row across both heatmaps is a clone; color indicates the proportion of the clone in each lineage in the respective wells. Example clones are shown on the right as red dots on SPRING plots. (**E** and **H**) Fate prediction from late isolated sisters is more accurate than early prediction for different machine-learning methods: NB, naïve Bayes; KNN, k-nearest neighbor; RF, random forest; LR, logistic regression; MLP, multilayer perceptron. Error bars indicate standard deviation across 100 partitions of the data into training and testing sets. (**I**) Split-well test for committed cells by sampling clones both on day 2 and in two separate wells on day 6. Clones emerging from pure multipotent states will show statistically independent fate outcomes in two wells (left), contrasting with committed clones (right). (**J**) scSeq SPRING plots showing early progenitors (day 2) colored by the fates of sisters isolated in separate wells (white dots indicate "mixed clones" with distinct fate outcomes). For each fate decision, the observed frequency of mixed clones falls short of that predicted for uncommitted progenitors, even for clusters most enriched for mixed clones (bottom panels). (**K** and **L**) Plot of predicted versus observed frequency of mixed clones. Points on the diagonal correspond to independent stochastic fate choice, points above the diagonal to asymmetric sister-cell fate, and points below the diagonal to fate priming or precommitment. For all fate choices studied, fate priming or precommitment is inferred.

by discrete, stepwise transitions in state and fate potential.

We interrogated the gene expression heterogeneity defining this continuum and its fate potential. The MPP ($CD34^+$) fraction of day 2 cells (Fig. 2E) contained several broad domains, including a restricted central domain of stem cell marker (Procr) expression; a wing expressing Gata2, an erythroid and stem cell marker; and an opposing wing expressing Flt3, indicative of lymphoid priming. Overlaying clonal outcomes (Fig. 2F) revealed regions of functional lineage priming consistent with these broad expression domains but further segregated into subdomains. Mk, Ba, Ma, and Eos potential were all restricted to the $Gata2^+$ region yet derived from separate subsets within this region. Testing for differential gene expression, we identified genes enriched within each subdomain of fate potential (Fig. 2G), revealing known markers and many that have not

been characterized in hematopoiesis [n = 447 (391 unique) differentially expressed genes at false discovery rate (FDR) = 0.05; table S3]. For example, Ikaros family zinc finger 2 (Ikzf2), a myeloid leukemia gene not previously associated with fate choice, was enriched in Eos and Ma progenitors but not Ba or Mk.

We similarly identified gene expression correlated with fate outcomes in less differentiated ST-HSCs and LT-HSCs transplanted into irradiated mice. As before, the cells spanned a continuous landscape with domains of primed gene expression, including a central domain of stem cell (Procr) and opposing wings of Gata2 and Flt3 expression (Fig. 2H) that correlated with output into the nine respective post-transplantation fates (Fig. 2I). Despite the less mature state of these cells, each fate outcome correlated with unique enriched genes before transplantation [Fig. 2J; n = 190 (173 unique) differentially expressed genes at FDR = 0.05; table S3], indicating specific priming at this early stage of differentiation. The differentially expressed genes represented a wide range of functional gene categories, from cell adhesion to chromatin regulation to intracellular and extracellular signaling, with cytokine signaling as the major enriched category ($p < 10^{-5}$; table S4). Gene-set enrichment analysis for each fate revealed terms associated with the fate's function, such as "lymphocyte activation" (p = 0.002 for T cell progenitors) and "response to bacterium" (p = 0.001 for Neu progenitors). Most of the top terms enriched in Er progenitors related to cell motility (8 of the top 10 terms; table S5), possibly indicating that these progenitors are primed to undergo cytoskeletal and niche rearrangements. We observed differences in clonal fate of phenotypically similar progenitors (day 2) in vivo compared with in vitro (fig. S6). Such environmental plasticity acts at subclonal resolution, as seen in an additional experiment by barcoding HSPCs and culturing them with different cytokines (n = 958 clones sampled between conditions; n = 1600 clones across time points within conditions; fig. S7, a to d). When split across cytokine conditions, sister cells showed consistent shifts of clone size and observed cell fate (fig. S7, e to g).

Overall, these observations support the view that functional lineage priming varies across a continuous hematopoietic progenitor landscape and covaries with the heterogeneous expression of genes, including transcription factors and a wide array of other functional gene categories. The observed clonal outcomes reflect both priming and environmental inputs.

How predictable is cell fate from gene expression?

Several factors influence the fate choice of a cell, including interactions with the environment, gene expression, chromatin state, and stochastic molecular events. scSeq provides only a limited view of cell state. Up to this point, we have considered the correlates of future fate choice revealed by this measurement. We now asked to what extent can fate be predicted from scSeq data?

To estimate the predictability of fate choice from gene expression, we considered the machine-learning task of predicting a cell's dominant fate outcome (Fig. 3, A and B) on the basis of its present scSeq profile (see materials and methods, section 9.1). We used two machine-learning methods: logistic regression and a neural network (multilayer perceptron). We applied these methods to several sets of genes, including all highly variable genes, genes that are differentially expressed between progenitors (table S3), and a genome-wide set of transcription factors (n =1811). Transcription factors were only marginally more informative than random size-matched gene sets (10% more informative in vitro; 3% more informative in vivo), whereas differentially expressed genes were substantially more informative (38% more informative in vitro; 20% more informative in vivo). Augmenting the differentially expressed genes with all highly variable genes, which increased the number of genes used by 12-fold in vitro and by 28-fold in vivo, did not significantly increase the accuracy (1% change in vivo, –4% change in vitro, $P > 0.05$). These results suggest that the predictive content of our gene expression measurements in HSPCs is almost entirely contained within several hundred differentially expressed genes, and only marginally enriched in transcription factors. The poor performance of transcription factors may be due to their low and noisy expression levels or to the comparable influence of other functional gene categories. These results were recapitulated when predicting the full distribution of fate outcomes rather than just the dominant one (fig. S8, g to j). Viewing predictive accuracy at the single-cell level revealed greater accuracy for increasingly mature cells (fig. S8, k to n; materials and methods, section 9.2). Across all conditions, the highest overall predictive accuracy from transcriptional state was 60% in vitro and 51% in vivo. These figures provide a lower bound for the cell-autonomous influence of transcriptional state on cell fate.

Functional purity of scSeq-defined cell states

Although fate prediction accuracy could be limited by stochastic fluctuations in cells or their environment, it is also possible that stable cellular properties influence fate choice but are not detected by scSeq. If such "hidden variables" (*4*) exist, then they would challenge the view that scSeq can define functionally pure populations. We tested for the presence of hidden variables by comparing "early" and "late" modes of cell fate prediction. If there were no hidden variables, then we reasoned that the information shared between separated sister cells could only decrease as time passes. Conversely, if there are stable properties that influence cell fate but are hidden from scSeq, then the mutual information between sisters could increase over time as these properties manifest in cell fate. This reasoning reflects a formal result known as data-processing inequality (*20*) (materials and methods, section 10.1).

To compare the accuracy of early versus late prediction, we applied a panel of machine-learning algorithms to guess the dominant fate of a clone using either the transcriptomes of its day 2 sisters (as in Fig. 3, A and B) or the transcriptomes of its sisters separated for 4 days in culture (n = 502 clones) or 1 week after transplantation (n = 69 clones) (Fig. 3, C to H). We found that late prediction was more informative for all algorithms tested (Fig. 3, E and H), with the most accurate algorithms achieving a late prediction accuracy of 76% in vitro and 70% in vivo compared with 60% and 52%, respectively, for early prediction.

These improvements in accuracy for late prediction reflect the high rate of concordance between sister-cell fates and hold true for clones of all potencies (Fig. 3D), consistent with recent observations of clonal fate restriction among HSPCs (*10*). Clones in separate wells produced identical combinations of fates 70% of the time compared with 22% by chance. One week after transplantation, sister cells in separate mice also showed highly concordant fate outcomes (Fig. 3G): Although they only shared the exact same combination of fates 29% of the time (compared with 10% by chance), they shared the same dominant fate 71% of the time (23% by chance). Together, these results imply that, both in culture and during transplantation, there are heritable properties of cell physiology that influence cell fate but are not evident in our scSeq measurements. We cannot tell whether information on cell fate is restricted simply because scSeq data are noisy, or if cell fate depends on cellular properties that are not reflected in the transcriptome, such as chromatin state, protein abundances, cell organization, or the microenvironment.

If scSeq states are not functionally pure, then phenotypically similar progenitors should be primed toward different fates. We tested this prediction by analyzing clones that were detected in three separate samples from our in vitro dataset: at day 2 and in two wells separated until day 6 (n = 408 clones; Fig. 3I). Without hidden variables, the two fates observed at day 6 should be statistically independent after conditioning on the day 2 state. In this case, the expected frequency of different fate outcome in the separate wells ("mixed clones") can be calculated (Fig. 3I, left; materials and methods, section 10.4). As a result of fate priming, however, we predicted that the frequency of mixed clones rooted in phenotypically similar day 2 cells would fall below

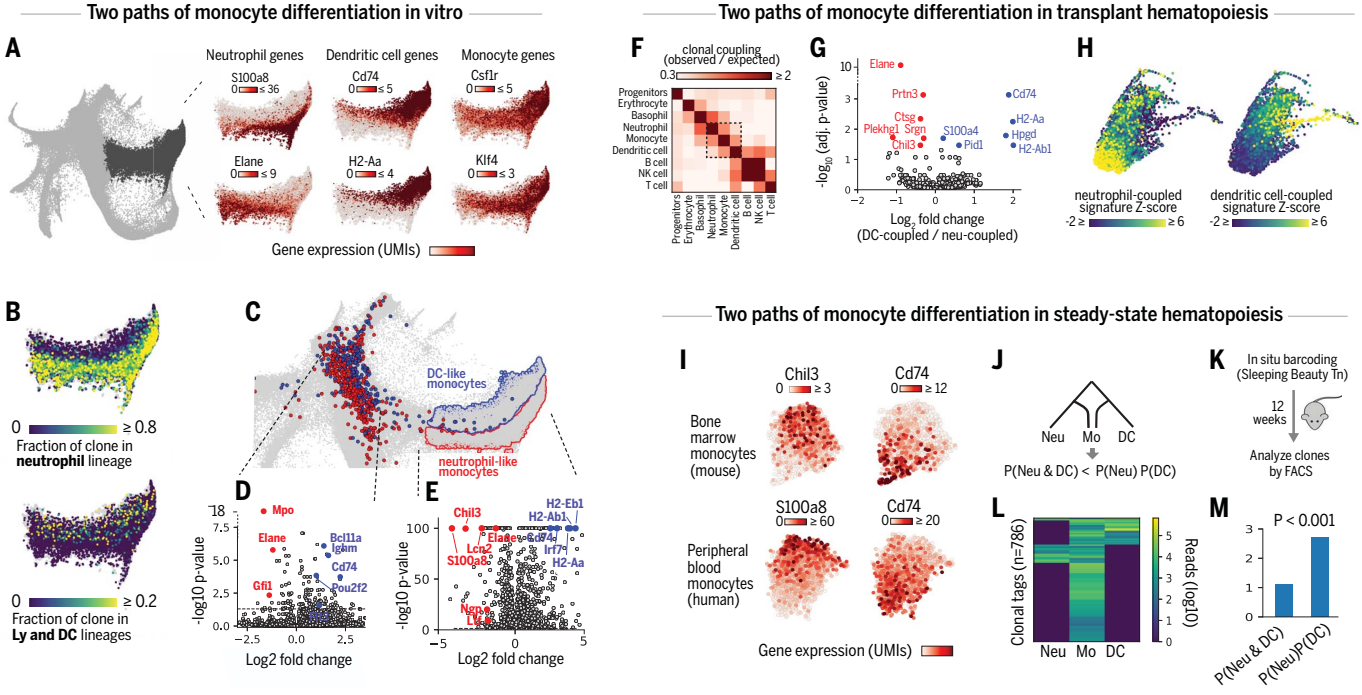


Fig. 4. Multiple paths of Mo differentiation. (**A**) Differentiating Mos show opposing expression of Neu and DC markers. Raw expression values are plotted with points ordered by expression level. (**B**) Mos segregate by proportions of Neu and DC sisters. Only Mos for which clonal data was available are shown. Plots show raw unsmoothed values from cells with clonal data. Points with the highest value are plotted on top. (**C**) Early (day 2) progenitors with sisters that differentiate into Neu-like or DC-like Mos occupy distinct transcriptional states. Plot is as in Fig. 2C. (**D** and **E**) Volcano plots identifying differentially expressed genes between the progenitors of (D) and mature (E) DC-like and Neu-like Mos. (**F**) Barcodes overlapping between cell types indicating Mo-DC and Mo-Neu coupling 1 week after transplantation. (**G**) Genes differentially expressed between Mos related to Neus or to DCs after transplantation. (**H**) Signature scores (average of *Z*-scored expression) shown on a SPRING plot of posttransplantation Mos. Points are ordered by expression level. (**I**) DC-to-Neu axis of gene expression persists in mature Mos, as seen by SPRING plots of scSeq data from Mos in mouse bone marrow (top) and human blood (bottom). (**J** to **M**) Clonal analysis of Mo differentiation in unperturbed hematopoiesis. (J) Under a model of two different Mo differentiation pathways, Neu-DC-Mo clones should be depleted relative to the null expectation. (K) Experimental schematic for barcoding mouse bone marrow in situ with clonal cell type composition assayed after a 12-week chase. (L) Number of cells in each type detected per clone (rows). (M) Observed versus independent expectation for Mo-Neu-DC clones is consistent with two Mo ontogenies.

this expectation. For each of three fate choices [Neu versus Mo, Neu/Mo versus Er/Mk/Ma/Ba, and Ly/DC versus all myeloid] and across different day 2 progenitor states, the proportion of mixed clones was below the expectation for pure bipotency (Fig. 3, J to L, and fig. S9, a and b). This analysis supports the previous conclusion that cell-autonomous fate biases can indeed coexist in the same measured scSeq state.

The above evidence for hidden variables suggests limits to the use of scSeq in building atlases that resolve the functional complexity of HSPCs. For years, cytometry [fluorescence-activated cell sorting (FACS)] has been used to examine the hematopoietic hierarchy with increasing precision, with the ultimate goal of defining functionally pure subsets of progenitors. Recent studies showing that many commonly used FACS gates are heterogeneous in fate and transcriptional state have raised the possibility that genome-wide assays such as scSeq might be required to achieve the necessary resolution. These results indicate that scSeq, although informative, may still be insufficient for defining functionally pure progenitor states.

Distinct routes of monocyte differentiation

Clonal analysis can reveal differentiation paths that may not be apparent by scSeq alone. In the data, Mos appeared to form a spectrum from Neu-like to DC-like, expressing alternatively Neutrophil elastase (Elane) and other Neu markers or major histocompatibility complex class II components (Cd74 and H2-Aa) (Fig. 4A). No similar overlap occurred with other cell types (fig. S10a). We investigated whether this phenotypic spectrum might result from distinct differentiation trajectories of Mos (*21*).

To determine Mo ontogenies, we scored their clonal relatedness with mature Neus and DCs. The Mos were not uniformly coupled to either cell type (Fig. 4B): those with increased expression of neutrophilic markers were clonally related to Neus (fig. S10b; $p < 10^{-7}$, Mann–Whitney *U* test), whereas those with DC-like gene expression were clonally related to DCs and lymphoid cells ($p < 10^{-17}$). We did not observe a comparable phenomenon for any other cell type in our data. Thus, only Mos appear to have a phenotypic spectrum that correlates with distinct clonal histories.

The distinct clonal origins of Mos suggested that they arise from progenitors with different fate potentials—and possibly different gene expression. To define their progenitors, we classified the differentiating Mos (4 to 6 days) as either DC-like or Neu-like (materials and methods, section 11.1) and then examined their early sisters (2 days). Indeed, the predecessors of DC-like and Neu-like Mos segregated by gene expression (Fig. 4, C and D), with respective expression of early DC and lymphoid markers (Flt3, Bcl11a, and Cd74) or early Neu markers (Elane, Mpo, and Gfi1; see table S6 for a full list of differentially expressed genes). These early differences were mostly distinct from those distinguishing mature (4 to 6 days) DC-like and Neu-like Mos (Fig. 4E and table S7). Our data therefore contain two different pathways of Mo differentiation with

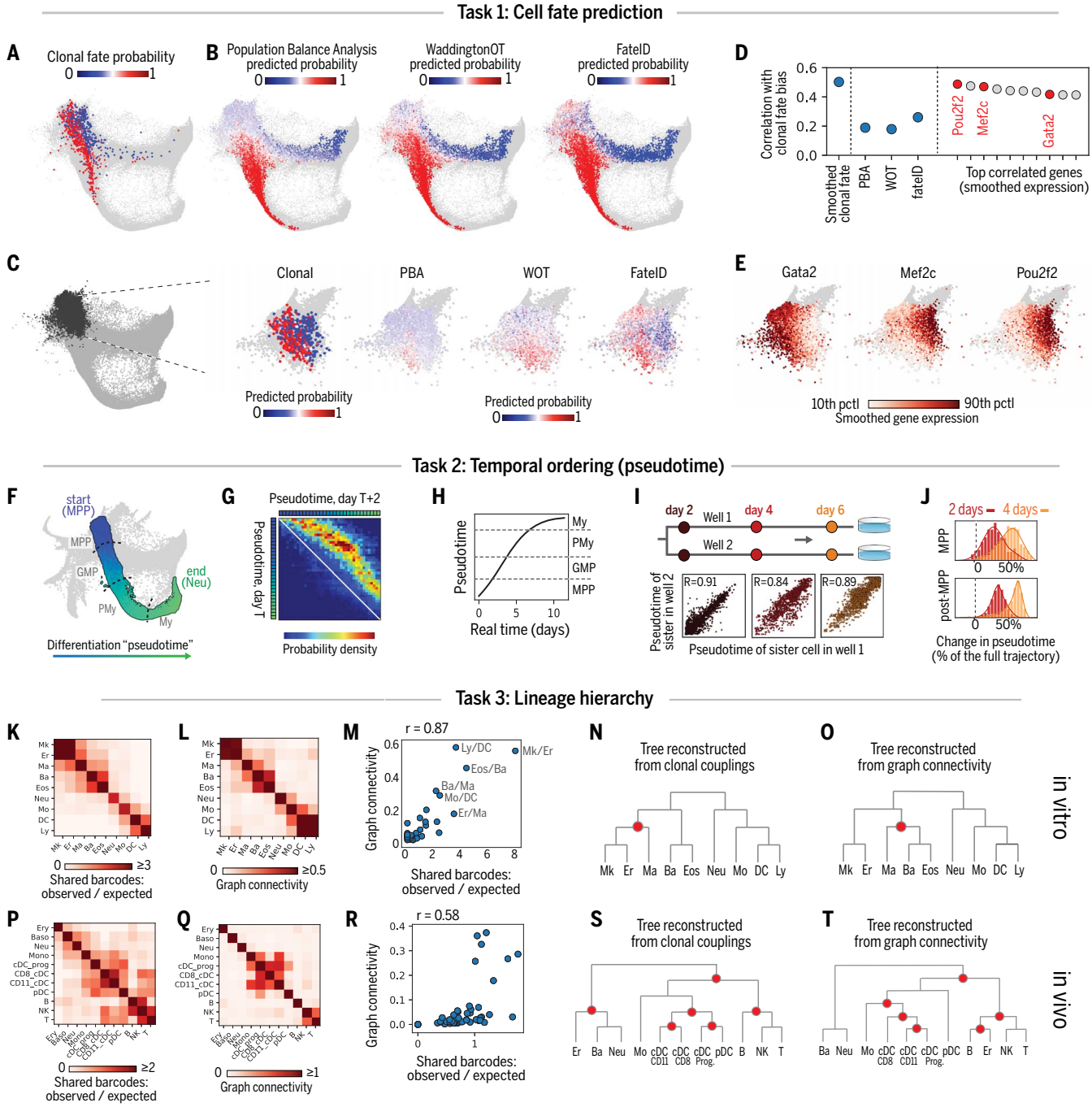


Fig. 5. Benchmark for dynamic inference from scSeq data. (**A**) SPRING plot of Neu and Mo differentiation, with progenitors (day 2) colored by the ratio of the Neu versus Mo fate of their sisters (days 4 to 6). (**B**) Algorithmic predictions of Neu versus Mo fate from transcription alone fail to recognize the early fate boundary revealed by clonal tracking. (**C**) Expanded view of early progenitors (thresholded by CD34 expression). Plots are as in (A) and (B). (**D**) Pearson correlation between future clonal fate outcomes of early progenitors and smoothed fate probabilities of held-out clonal data, output of algorithmic predictions, and expression of top 10 most correlated genes (red, transcription factors). Held-out data set the upper bound on accuracy of fate prediction algorithms. (**E**) Expression of fate-correlated transcription factors in CD34^{+} progenitors. Points are ordered by expression level. (**F**) "Pseudotime" ordering of Neu differentiation. Dotted lines represent the approximate boundaries in gene expression associated with canonical stages (PMy, promyelocyte; My, myelocyte). (**G**) Joint distribution of pseudotime of sister cells separated in time by 2 days reveals a consistent forward shift across the trajectory. (**H**) Pseudotime progression as a function of real time obtained from integration of pseudotime velocity from (G). (**I**) Pseudotime remains correlated for sister cells cultured in separate wells. (**J**) Distributions of pseudotime changes showing greater variability in MPPs compared with later stages (red, days 2 to 4; orange, days 2 to 6). (**K**) Clonal overlap between cell types in culture. The number of shared barcodes between pairs is normalized by expectation if clonal membership is shuffled. (**L**) State proximity for cell types in culture, represented by graph diffusion distance (connectivity) in a high-dimensional KNN graph of all data from Fig. 1E. (**M**) Clonal overlap across all pairs of lineages correlates with state proximity. (**N** and **O**) Inferred differentiation hierarchies assembled by iteratively joining cell types on the basis of clonal or state distances. Red dots indicate the sole discrepancy between the hierarchies. (**P** to **T**) As in (K) to (O) repeated for cells after transplantation, showing increased discrepancies between clonal and state-based hierarchies.

distinct clonal relationships and gene expression dynamics.

These results are consistent with a recent finding that immunophenotypically defined Mo-dendritic progenitors (MDPs) and granulocyte-Mo progenitors (GMPs) give rise to Mos with DC-like and Neu-like characteristics, respectively (*21*). To test whether our observations represent MDP or GMP outputs, we performed scSeq on fresh MDPs and GMPs sorted from adult mouse bone marrow and found that they colocalized with the day 2 progenitors of DC-like and Neu-like Mos. Similarly, scSeq analysis of MDPs and GMPs cultured for 4 days in vitro colocalized with mature DC-like and Neu-like Mos (fig. S10c). Thus, the DC-like and Neu-like trajectories observed here likely represent MDP and GMP pathways of Mo differentiation and clarify the location of these states in a gene expression continuum.

Several lines of evidence support the existence of distinct Mo-Neu and Mo-DC clonal couplings in vivo, not just in culture: (i) clonal and gene expression relationships after transplantation; (ii) persistent heterogeneity in freshly isolated mouse and human Mos; and (iii) results from nonperturbative in vivo clonal analysis. We present these results in turn.

First, 1 week after transplantation, Mos showed distinct clonal relationships to Neus and DCs (Fig. 4F). As was the case in vitro, the DC-related Mos were enriched for DC marker genes, whereas Neu-related Mos were enriched for Neu markers (Fig. 4, G and H). Second, we analyzed classical Mos (fig. S10g) and human peripheral blood Mos (fig. S10h) by scSeq. Principal component analysis showed that in both cases there was a spectrum of Neu-like to DC-like gene expression (see table S8 for differentially expressed genes), which was also evident in the expression of marker genes (Fig. 4I). This analysis agrees with earlier observations (*21*). Third, in native hematopoiesis, we examined the clonal cooccurrence of Mos with DCs and Neus after genetically barcoding HPC clones in a nonperturbative manner using a transposase-based strategy (*22*) (materials and methods, section 12.3). If Mo heterogeneity correlates with distinct clonal coupling to Neus versus DCs, then we would expect an anticorrelation between Neu and DC relatedness among Mos (Fig. 4J). After a 12-week chase (Fig. 4K), we indeed found significantly fewer Neu-Mo-DC tags than would be expected if clonal cooccurrence were independent (2.5-fold reduction; $p < 0.001$ by binomial test of proportion; Fig. 4, L and M). Overall, our results support the existence of multiple Mo ontogenies in native hematopoiesis as well as in culture and during transplantation.

Benchmark for fate prediction in hematopoiesis

To understand hematopoietic fate control, we and others have been interested in developing data-driven models of gene expression dynamics constrained by scSeq data (*3*, *4*, *5*, *7*, *23*). Computational models could identify cellular components driving fate choice and the sequence of gene expression changes that accompany cell maturation. Because of the lack of ground truth data, existing methods have been difficult to compare and validate. Here, we asked whether common approaches for modeling cell-state dynamics are consistent with our clonal tracking data.

scSeq-based models do not fully predict fate choice

We first asked how well existing computational models using only scSeq data predict cell fate probabilities. We tested three recent approaches, population balance analysis (*4*), WaddingtonOT (*5*), and FateID (*7*), for their ability to predict the fate of a cell choosing between Neu and Mo fates in culture. We calculated for each cell at day 2 the fraction of its clonal relatives that became a Neu or a Mo (Fig. 5A) and then attempted to predict this fraction from transcriptomes alone (Fig. 5B; materials and methods, sections 13.2 to 13.4). All three methods were broadly consistent with clonal fate bias as cells began to mature, but in the early progenitor ($Cd34^+$) region, clonal tracking revealed a bifurcation of Mo and Neu potential that was generally not detected by the prediction algorithms, although FateID performed slightly better (Fig. 5, C and D; $R < 0.26$ for all methods). All fell considerably below fate predictions obtained from held-out clonal data ($R = 0.5$; materials and methods, section 13.5; correlation is low overall because of noise in the fate outcomes of single cells). These results show that in the absence of lineage information, computational methods may misidentify fate decision boundaries. It is therefore important that when genes are ranked by their ability to predict cell fate bias, the top 10 genes easily outperformed the prediction algorithms (Fig. 5D), including known fate regulators such as Gata2 and Mef2c (Fig. 5, D and E). The selection of the correct genes to use for prediction, however, required clonal information. These results provide a framework for comparing computational models of differentiation and may serve as a useful benchmark for improving them.

Temporal progression is captured by pseudotime

A common goal of scSeq is to order gene expression along dynamic trajectories by defining a "pseudotime" coordinate that orders transcriptomes (*24*). At present, it is unknown how single cells traverse these trajectories, including whether they progress at different rates or even reverse their dynamics (*4*). Focusing on Neu differentiation as a test case, we asked how well pseudotime describes the kinetics of differentiation as revealed by clonal tracking. We ordered cells from MPPs to GMPs to promyelocytes to myelocytes ($n = 63{,}149$ cells; Fig. 5F; fig. S11a; materials and methods, section 14.2) and compared the pseudotemporal progression of clones sampled at 2 consecutive days (Fig. 5G). This analysis showed a consistent forward velocity along differentiation pseudotime. By integrating the velocity across the trajectory, we were able to calculate pseudotime progression as a function of real time for a typical cell (Fig. 5H; materials and methods, section 14.3). The time for an MPP to differentiate into a myelocyte was 10 days, consistent with previous results (*25*). Pseudotime analysis of sister cells differentiated in separate wells also showed a consistent pace of differentiation both shortly after cell division (day 2) and 4 days later ($R \geq 0.89$; Fig. 5I). Pseudotime velocity was most variable among MPPs (Fig. 5J), which could be explained by cells remaining in the MPP state for a variable duration before initiating Neu differentiation. These results support the use of pseudotime methods for mapping differentiation progression.

Agreement of state and clonal differentiation hierarchies

For cells undergoing multilineage fate choice, scSeq has been used to estimate lineage hierarchy on the basis of the assumption that cell types with transcriptionally similar differentiation pathways are clonally related (*3*–*5*, *7*). However, this assumption may not always hold: Similar end states could also arise from nonoverlapping clones (*26*), and distant end states could share lineage through asymmetric division.

To compare fate hierarchies constructed using lineage and state information, for each pair of differentiated states, we quantified the number of shared clones as well as the similarity of cell states for each pair of differentiated fates both in vivo and in culture (Fig. 5, K, L, P, and Q; materials and methods, sections 15.1 and 15.2). We found that measures of state distance and clonal coupling are closely correlated in vitro ($r = 0.93$, $p < 10^{-35}$; Fig. 5M). When we constructed candidate cell-type hierarchies from state distance and clonal distance, respectively (Fig. 5, N and O), they were almost identical, with only one difference in the differentiation path assigned to Mas. These results held for a broad range of parameters and for different distance metrics (fig. S12, a to h). In vivo, however, the same analysis revealed a weaker correlation between state and fate distance ($r = 0.58$; $p = 0.065$; Fig. 5R), with considerable differences between the resulting cell-type hierarchies (Fig. 5, S and T). Several factors might explain the weaker relationship between state and fate hierarchy in vivo, such as the longer interval between samples (1 week, compared with every 2 days in vitro) or the complex differentiation environment. These results

suggest a set of experimental parameters—operant in our in vitro experiment—that may be favorable for inferring clonal relationships from gene expression topology: dense sampling over time, uniformity of the differentiation environment, and a spectrum of the maturity in the initially barcoded cells.

DISCUSSION

LARRY defines a scSeq-compatible lineage-tracing approach that links cell states to clonal fates simultaneously from multiple initial conditions without the need to target each specific progenitor state. The strategy differs from CRISPR-based lineage-tracing approaches (*27*, *28*) in that it links states across time, not only at a single end point. LARRY is simple to use; unlike CRISPR-based approaches, it does not require lineage tree inference to establish sister-cell relationships, it exhibits very low single-cell barcode dropout rates, and it does not require delivering multiple components. As with CRISPR-based approaches, the method cannot study processes faster than one cell cycle. It is currently restricted to culture or transplantation assays. However, within this constraint, the approach allows correlating early gene expression with fate in an unbiased manner, avoiding boundaries imposed by a particular choice of reporter gene or by cell-sorting criteria. We demonstrated that this strategy can be simply extended to paired perturbation experiments that compare sister cells treated in different conditions.

In hematopoiesis, a long-term goal has been to define a complete atlas of progenitor cell states and their fate potentials as a basis for understanding fate control. Here, we confirmed that functional lineage priming in MPPs is associated with low-level expression of lineage-affiliated genes, including transcription factors and a wide array of other functional gene categories, and that cells differentiate through a continuous, structured fate hierarchy that differs from classical tree-like depictions of hematopoiesis in its clonal structure. We additionally found evidence for a revised ontogeny of Mos (*21*) in culture, transplantation, and native hematopoiesis. In addition to locating fate bias on a single-cell landscape, our results revealed the limits of scSeq to distinguish functionally heterogeneous states by showing that transcriptionally similar cells can have a cell-autonomous bias toward different fate choices. The molecular factors distinguishing these cells may be undersampled mRNA or heritable cellular properties such as chromatin state that are hidden from scSeq but manifest in the fate of isolated sister cells. Our results thus argue for looking beyond scSeq alone in defining cellular maps of adult and developing tissues. Coupling cell state and fate readouts in different tissues will deepen our understanding of stem cell behaviors in tissue development and homeostatic physiology.

REFERENCES AND NOTES

1. P. Jensen, S. M. Dymecki, Essentials of recombinase-based genetic fate mapping in mice. *Methods Mol. Biol.* **1092**, 437–454 (2014). doi: 10.1007/978-1-60327-292-6_26; pmid: 24318835
2. M. B. Woodworth, K. M. Girskis, C. A. Walsh, Building a lineage from single cells: Genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017). doi: 10.1038/nrg.2016.159; pmid: 28111472
3. C. A. Herring, B. Chen, E. T. McKinley, K. S. Lau, Single-cell computational strategies for lineage reconstruction in tissue systems. *Cell. Mol. Gastroenterol. Hepatol.* **5**, 539–548 (2018). doi: 10.1016/j.jcmgh.2018.01.023; pmid: 29713661
4. C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, A. M. Klein, Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2467–E2476 (2018). doi: 10.1073/pnas.1714723115; pmid: 29463712
5. G. Schiebinger *et al.*, Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 1517 (2019). doi: 10.1016/j.cell.2019.02.026; pmid: 30849376
6. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi: 10.1038/s41586-018-0414-6; pmid: 30089906
7. J. S. Herman, D. Sagar, D. Grün, FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat. Methods* **15**, 379–386 (2018). doi: 10.1038/nmeth.4662; pmid: 29630061
8. K. Akashi, D. Traver, T. Miyamoto, I. L. Weissman, A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000). doi: 10.1038/35004599; pmid: 10724173
9. S. Nestorowa *et al.*, A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31 (2016). doi: 10.1182/blood-2016-05-716480; pmid: 27365425
10. L. Perié, K. R. Duffy, L. Kok, R. J. de Boer, T. N. Schumacher, The branching point in erythro-myeloid differentiation. *Cell* **163**, 1655–1662 (2015). doi: 10.1016/j.cell.2015.11.059; pmid: 26687356
11. F. Notta *et al.*, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016). doi: 10.1126/science.aab2116; pmid: 26541609
12. L. Kester, A. van Oudenaarden, Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* **23**, 166–179 (2018). doi: 10.1016/j.stem.2018.04.014; pmid: 29754780
13. L. Tian *et al.*, SIS-seq, a molecular 'time machine', connects single cell fate with gene programs. bioRxiv 403113 [Preprint]. 29 August 2018. https://doi.org/10.1101/403113.
14. D. T. Montoro *et al.*, A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018). doi: 10.1038/s41586-018-0393-7; pmid: 30069044
15. R. Lu, N. F. Neff, S. R. Quake, I. L. Weissman, Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928–933 (2011). doi: 10.1038/nbt.1977; pmid: 21964413
16. D. S. Lin *et al.*, DiSNE movie visualization and assessment of clonal kinetics reveal multiple trajectories of dendritic cell development. *Cell Rep.* **22**, 2557–2566 (2018). doi: 10.1016/j.celrep.2018.02.046; pmid: 29514085
17. B. A. Biddy *et al.*, Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018). doi: 10.1038/s41586-018-0744-4; pmid: 30518857
18. C. Weinreb, S. Wolock, A. M. Klein, SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018). doi: 10.1093/bioinformatics/btx792; pmid: 29228172
19. L. Velten *et al.*, Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017). doi: 10.1038/ncb3493; pmid: 28319093
20. J. B. Kinney, G. S. Atwal, Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3354–3359 (2014). doi: 10.1073/pnas.1309933111; pmid: 24550517
21. A. Yáñez *et al.*, Granulocyte-monocyte progenitors and monocyte-dendritic cell progenitors independently produce functionally distinct monocytes. *Immunity* **47**, 890–902.e4 (2017). doi: 10.1016/j.immuni.2017.10.021; pmid: 29166589
22. J. Sun *et al.*, Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014). doi: 10.1038/nature13824; pmid: 25296256
23. B. K. Tusi *et al.*, Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018). doi: 10.1038/nature25741; pmid: 29466336
24. C. Trapnell *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). doi: 10.1038/nbt.2859; pmid: 24658644
25. D. Gupta, H. P. Shah, K. Malu, N. Berliner, P. Gaines, Differentiation and characterization of myeloid cells. *Curr. Protoc. Immunol.* **104**, 1, 28 (2014). doi: 10.1002/0471142735.im22f05s104; pmid: 24510620
26. D. E. Wagner *et al.*, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018). doi: 10.1126/science.aar4362; pmid: 29700229
27. M. M. Chan *et al.*, Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019). doi: 10.1038/s41586-019-1184-5; pmid: 31086336
28. B. Spanjaard *et al.*, Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018). doi: 10.1038/nbt.4124; pmid: 29644996

ACKNOWLEDGMENTS

We thank the Single Cell Core Facility at Harvard Medical School for inDrop reagents, the Bauer Core Facility for sequencing, B. Gottgens for discussions and comments on the manuscript, and K. Kawaguchi for mentorship in experiments and analysis. **Funding:** A.M.K. and C.W. were supported by NIH grant nos. R33CA212697-01 and 1R01HL14102-01, Harvard Stem Cell Institute Blood Program Pilot grant no. DP-0174-18-00, and Chan-Zuckerberg Initiative grant no. 2018-182714. A.R.-F was supported by a Merck Fellowship from the Life Sciences Research Foundation, an EMBO Long-Term Fellowship (ALTF 675-2015), an American Society of Hematology Scholar Award, a Leukemia & Lymphoma Society Career Development Program Award (3391-19), and NIH grant no. K99HL146983. F.D.C. was supported by NIH grant nos. HL128850-01A1 and P01HL13147. F.D.C. is a scholar of the Howard Hughes Medical Institute and the Leukemia & Lymphoma Society. **Author contributions:** All authors designed the experiments. C.W. and A.R.-F. performed the experiments. C.W. carried out the computational analysis. C.W. and A.M.K. wrote the paper. A.M.K. and F.D.C. jointly supervised the work. **Competing interests:** A.M.K. is a founder of 1Cell-Bio, Ltd. **Data and materials availability:** Raw gene expression data and processed counts are available on GEO, accession no. GSE140802. Further data are available at github.com/AllonKleinLab/paper-data.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6479/eaaw3381/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S12
References (*29*–*35*)

10 December 2018; accepted 14 January 2020
Published online 23 January 2020
10.1126/science.aaw3381

Science

Lineage tracing on transcriptional landscapes links state to fate during differentiation

Caleb Weinreb, Alejo Rodriguez-Fraticelli, Fernando D. Camargo and Allon M. Klein

Science **367** (6479), eaaw3381.
DOI: 10.1126/science.aaw3381originally published online January 23, 2020

Mapping cell fate during hematopoiesis

Biologists have long attempted to understand how stem and progenitor cells in regenerating and embryonic tissues differentiate into mature cell types. Through the use of recent technical advances to sequence the genes expressed in thousands of individual cells, differentiation mechanisms are being revealed. Weinreb *et al.* extended these methods to track clones of cells (cell families) across time. Their approach reveals differences in cellular gene expression as cells progress through hematopoiesis, which is the process of blood production. Using machine learning, they tested h ow well gene expression measurements account for the choices that cells make. This work reveals that a considerable gap still exists in understanding differentiation mechanisms, and future methods are needed to fully understand—and ultimately control—cell differentiation.

Science, this issue p. eaaw3381