

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†‡}

¹Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. ²Stem Cell Program, Boston Children's Hospital, Boston, MA 02115, USA. ³Department 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

A challenge in biology is to associate molecular differences among progenitor cells with their capacity to generate mature cell types. Here, we use expressed DNA barcodes to clonally trace transcriptomes over time and applied this to study fate determination in hematopoiesis. We identify states of primed fate potential and locate them on a continuous transcriptional landscape. We identify two routes of monocyte differentiation that leave an imprint on mature cells. Yet analysis of sister cells also reveals cells to have intrinsic fate biases not detectable by single-cell RNA sequencing. Finally, we benchmark 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, where 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) opened up a complementary approach to understand developmental relationships. scSeq captures mature cell types alongside 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 carried out without prior genetic manipulation, and without being limited by the specificity of transgene expression within the progenitor cell pool (2).

Neither state or lineage mapping alone, however, provide a complete view of 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 does 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.

In this paper 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). But 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). 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

A 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 the remainder later (13). This approach provides data for three types of clonal relationships (Fig. 1A): (1) sister cells in the earliest time point may be captured after 1 or 2 rounds of division; (2) clones observed at both early and later time points allow comparing the state of an early cell to the fate outcomes of its sisters; (3) sampling differentiated cells at later time points 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' UTR of an enhanced green fluorescent protein (*eGFP*) transgene under control of a ubiquitous EF1 α 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 $\sim 0.5 \times 10^6$ barcodes, sufficient to label 5000 cells in an experiment with <1% barcode overlap between clones (see materials and methods, section 2.3, for estimate of diversity). We refer to the barcoding construct as LARRY (Lineage And RNA Recovery).

We tested LARRY on mouse embryonic stem (ES) 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 to cells transplanted in vivo. For in vitro analysis, we isolated a broad class of oligo-potent (Lin-Sca-Kit+) and multipotent progenitors (Lin-Sca1+Kit+ or LSK) cells (fig. S2, a and b) and plated them in media chosen to support broad multi-lineage differentiation (see materials and methods). Following barcode transduction, cells were cultured for two 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 re-plated and then sampled after two days (30% of cells) and four days (remaining cells) (Fig. 1C). For transplantation, Lin-Sca(hi)Kit+ cells, consisting of mostly short-term and long-term hematopoietic stem cells (HSCs) (fig. S2, a and b), were barcoded and

placed in culture. After two days, 40% were profiled by scSeq, with the remainder transplanted into ten sublethally irradiated host mice (10) and recovered for scSeq one and two 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 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 1,816 and 817 clones in total spanning early and late timepoints (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), dendritic cells (plasmacytoid pDC; Ccr7+ migratory migDC) and lymphoid precursors (Ly). On this landscape clones exhibited a range of behaviors including uni-lineage and multi-lineage 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 neutrophil 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+, migDC and pDC, as well as Ly6C+ classical and Ly6C- non-classical monocytes (Fig. 1J and fig. S3). We did not detect megakaryocytes, possibly because they did not survive bone marrow harvest, flow-sorting and single-cell encapsulation intact. Therefore, with these experiments we simultaneously captured 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. Yet 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 a majority (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 to 80% for random cell pairs). We tested and ruled out that similar sister pairs are technical co-encapsulation artifacts (fig. S4e). These tests justified approximating single-cell trajectories by clonal trajectories, though with some loss in resolution of fate boundaries expected due to $\sim 10\%$ diverged sister pairs.

Beginning with the in vitro data, we recorded the clonal

fates of each day 2 cell. Visualizing cells from uni-lineage clones revealed well-delineated domains of fate potential (Fig. 2C). Where the progenitors for different fates overlapped, we observed bi-potent or oligopotent clones, indicating the location of fate commitment boundaries (Fig. 2D). The true number of multipotent clones is likely underestimated in our data, since some clonal fates were likely missed due to under-sampling (fig. S5) and cell commitment prior to 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 by discrete, step-wise transitions in state and fate potential.

We interrogated the gene expression heterogeneity defining this continuum and its fate potential. The multipotent progenitor (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. Megakaryocyte, basophil, mast cell and eosinophil 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 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 eosinophil and mast cell progenitors, but not basophil or megakaryocyte.

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-transplant fates (Fig. 2I). Despite the less mature state of these cells, each fate outcome correlated with unique enriched genes prior to 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 as well as intra- and extra-cellular 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 neutrophil progenitors). The majority of top terms enriched in erythrocyte progenitors related to cell motility (8 out of the top 10 terms; table S5), possibly indicating 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 to in vitro (fig. S6). Such environmental plasticity acts at sub-clonal resolution, as seen by barcoding HSPCs and culturing them in different cytokines (n=958 clones sampled between conditions; n=1,600 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 impinge on the fate choice of a cell, including interactions with the environment, gene expression, chromatin state, and stochastic molecular events. Single-cell RNA-seq provides only a limited view of cell state. So far, we have considered the correlates of future fate choice revealed by this measurement. We now ask: 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) based on its present scSeq profile (materials and methods, section 9.1). We used two machine learning methods: logistic regression and a neural network [multi-layer 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 28-fold in vivo,

did not lead to a significant change in accuracy (1% change in vivo, -4% change in vitro). 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 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, 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, 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 the Data Processing Inequality (20) (materials and methods, section 10.1).

To compare the accuracy of early vs 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 four days in culture (n=502 clones) or one week post-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 late prediction accuracy of 76% in vitro and 70% in vivo, compared to 60% and 52% respectively for early prediction.

These improvements in accuracy for late prediction reflect the high rate of concordance between sisters cell fates, and hold 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 to 22% by chance. One week

post-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 to 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 because 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 this expectation. For each of three fate choices (Neu vs. Mo, [Neu/Mo] vs. [Er/Mk/Ma/Ba], and [Ly/DC] vs. [all myeloid]), and across different day 2 progenitor states, the proportion of mixed clones was significantly below the expectation for pure bi-potency (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 (FACS) has been used to dissect 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, while 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, monocytes appeared to form a spectrum from neutrophil-like to DC-like, expressing alternatively neutrophil elastase (Elane) and other neutrophil markers, or MHC class II components (Cd74 and H2-

Aa) (Fig. 4A). No similar overlap occurs with other cell types (fig. S10a). We investigated whether this phenotypic spectrum might result from distinct differentiation trajectories of monocytes (21).

To determine monocyte ontogenies, we scored their clonal relatedness with mature neutrophils and DCs. The monocytes were not uniformly coupled to either cell type (Fig. 4B): those with increased expression of neutrophilic markers were clonally related to neutrophils (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, monocytes appear unique in showing a phenotypic spectrum that correlated with distinct clonal histories.

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

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

Several lines of evidence support the existence of distinct monocyte-neutrophil and monocyte-DC clonal couplings in vivo, and not only in culture: (i) clonal and gene expression relationships following transplantation; (ii) persistent heterogeneity in freshly-isolated mouse and human monocytes;

and (iii) results from non-perturbative in-vivo clonal analysis. We present these results in turn.

First, one week after transplantation, monocytes showed distinct clonal relationships to neutrophils and DCs (Fig. 4F). As in vitro, the DC-related monocytes were enriched for DC marker genes, whereas neutrophil-related monocytes were enriched for neutrophil markers (Fig. 4, G and H). Second, we analyzed classical monocytes (fig. S10g) and human peripheral blood monocytes (fig. S10h) by scSeq. Principal component analysis shows that in both cases a spectrum exists of neutrophil-like to DC-like gene expression (see table S8 for differentially expressed genes), which is 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 co-occurrence of monocytes with DCs and neutrophils after genetically barcoding HPC clones in a non-perturbative manner using a transposase-based strategy (22) (materials and methods, section 12.3). If monocyte heterogeneity correlates with distinct clonal coupling to neutrophils versus DCs, we would expect an anti-correlation between neutrophil and DC relatedness among monocytes (Fig. 4J). After a 12 week chase (Fig. 4K), we indeed found significantly fewer neutrophil-monocyte-DC tags than would be expected if clonal co-occurrence 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 monocyte ontogenies in native hematopoiesis as well as in culture and during transplantation.

A 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. Due to a lack of ground truth data, existing methods have been hard 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 (PBA) (4), WaddingtonOT (WOT) (5), and FateID (7) for their ability to predict the fate of a cell choosing between neutrophil and monocyte fates in culture. We calculated for each cell at day 2 the fraction of its clonal relatives that became a neutrophil or a monocyte (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 monocyte and neutrophil 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 mis-identify fate decision boundaries. It is therefore significant that when genes are ranked by their ability to predict cell fate bias, the top ten 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 neutrophil 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 (PMy), to myelocytes (My) ($n=63,149$ cells; Fig. 5F; fig. S11a; materials and methods, section 14.2), and compared the pseudotemporal progression of clones sampled at two 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 prior literature (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 remaining so four 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 neutrophil differentiation. These results support the use of pseudotime methods for mapping differentiation progression.

Agreement of state and clonal differentiation hierarchies

For cells undergoing multi-lineage fate choice, scSeq has been

used to estimate lineage hierarchy based on the assumption that cell types with transcriptionally similar differentiation pathways are clonally related (3-5, 7). Yet this assumption may not always hold: similar end states could also arise from non-overlapping 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 mast cells. 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 to 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 needing 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 because temporal sampling disrupts spatial organization. Yet 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 states.

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 associates with low-level expression of lineage-affiliated genes, including transcription factors and a wide array of other functional gene categories, and that cells differentiate via 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 monocytes (2I) 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 cell-autonomous bias toward different fate choices. The molecular factors distinguishing these cells may be under-sampled 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 Medline
2. M. B. Woodworth, K. M. Girsakis, 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 Medline
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 Medline
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 Medline
5. G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, E. S. Lander, 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 Medline
6. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastrioti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Born, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi:10.1038/s41586-018-0414-6 Medline
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 Medline
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 Medline
9. S. Nestorowa, F. K. Hamey, B. Pijuan Sala, E. Diamanti, M. Shepherd, E. Laurenti, N. K. Wilson, D. G. Kent, B. Göttgens, 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 Medline
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 Medline
11. F. Notta, S. Zandi, N. Takayama, S. Dobson, O. I. Gan, G. Wilson, K. B. Kaufmann, J. McLeod, E. Laurenti, C. F. Dunant, J. D. McPherson, L. D. Stein, Y. Dror, J. E. Dick, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016). doi:10.1126/science.aab2116 Medline
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 Medline
13. L. Tian, J. Schreuder, D. Zalcenstein, J. Tran, N. Kocovski, S. Su, P. Diakumis, M. Bahlo, T. Sargeant, P. D. Hodgkin, M. E. Ritchie, S. H. Naik, 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, A. L. Haber, M. Biton, V. Vinarsky, B. Lin, S. E. Birket, F. Yuan, S. Chen, H. M. Leung, J. Villoria, N. Rogel, G. Burgin, A. M. Tsankov, A. Waghray, M. Slyper, J. Waldman, L. Nguyen, D. Dionne, O. Rozenblatt-Rosen, P. R. Tata, H. Mou, M. Shivaraju, H. Bihler, M. Mense, G. J. Tearney, S. M. Rowe, J. F. Engelhardt, A. Regev, J. Rajagopal, A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018). doi:10.1038/s41586-018-0393-7 Medline
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 Medline
16. D. S. Lin, A. Kan, J. Gao, E. J. Crampin, P. D. Hodgkin, S. H. Naik, 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 Medline
17. B. A. Bidy, W. Kong, K. Kamimoto, C. Guo, S. E. Wayne, T. Sun, S. A. Morris, Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018). doi:10.1038/s41586-018-0744-4 Medline
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 Medline
19. L. Velten, S. F. Haas, S. Raffel, S. Blaszkiewicz, S. Islam, B. P. Hennig, C. Kirche, C. Lutz, E. C. Buss, D. Nowak, T. Boch, W.-K. Hofmann, A. D. Ho, W. Huber, A. Trumpp, M. A. G. Essers, L. M. Steinmetz, Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017). doi:10.1038/ncb3493 Medline
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 Medline
21. A. Yáñez, S. G. Coetzee, A. Olsson, D. E. Muench, B. P. Berman, D. J. Hazelett, N. Salomonis, H. L. Grimes, H. S. Goodridge, 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 Medline
22. J. Sun, A. Ramos, B. Chapman, J. B. Johnnidis, L. Le, Y.-J. Ho, A. Klein, O. Hofmann, F. D. Camargo, Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014). doi:10.1038/nature13824 Medline
23. B. K. Tusi, S. L. Wolock, C. Weinreb, Y. Hwang, D. Hidalgo, R. Zilionis, A. Waisman, J. R. Huh, A. M. Klein, M. Socolovsky, Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018). doi:10.1038/nature25741 Medline
24. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, 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 Medline
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 Medline
26. D. E. Wagner, C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason, A. M. Klein, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018). doi:10.1126/science.aar4362 Medline
27. M. M. Chan, Z. D. Smith, S. Grosswendt, H. Kretzmer, T. M. Norman, B. Adamson,

- M. Jost, J. J. Quinn, D. Yang, M. G. Jones, A. Khodaverdian, N. Yosef, A. Meissner, J. S. Weissman, Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019). [doi:10.1038/s41586-019-1184-5](https://doi.org/10.1038/s41586-019-1184-5) [Medline](#)
28. B. Spanjaard, B. Hu, N. Mitic, P. Olivares-Chauvet, S. Janjuha, N. Ninov, J. P. Junker, Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018). [doi:10.1038/nbt.4124](https://doi.org/10.1038/nbt.4124) [Medline](#)
29. S. H. Naik, T. N. Schumacher, L. Perié, Cellular barcoding: A technical appraisal. *Exp. Hematol.* **42**, 598–608 (2014). [doi:10.1016/j.exphem.2014.05.003](https://doi.org/10.1016/j.exphem.2014.05.003) [Medline](#)
30. A. Gerrits, B. Dykstra, O. J. Kalmykova, K. Klauke, E. Verovskaya, M. J. C. Broekhuis, G. de Haan, L. V. Bystrykh, Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010). [doi:10.1182/blood-2009-06-229757](https://doi.org/10.1182/blood-2009-06-229757) [Medline](#)
31. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015). [doi:10.1016/j.cell.2015.04.044](https://doi.org/10.1016/j.cell.2015.04.044) [Medline](#)
32. L. W. Plasschaert, R. Žilionis, R. Choo-Wing, V. Savova, J. Knehr, G. Roma, A. M. Klein, A. B. Jaffe, A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018). [doi:10.1038/s41586-018-0394-6](https://doi.org/10.1038/s41586-018-0394-6) [Medline](#)
33. H. Mi, A. Muruganujan, P. D. Thomas, PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41** (D1), D377–D386 (2013). [doi:10.1093/nar/gks1118](https://doi.org/10.1093/nar/gks1118) [Medline](#)
34. M. Kanamori, H. Konno, N. Osato, J. Kawai, Y. Hayashizaki, H. Suzuki, A genome-wide and nonredundant mouse transcription factor database. *Biochem. Biophys. Res. Commun.* **322**, 787–793 (2004). [doi:10.1016/j.bbrc.2004.07.179](https://doi.org/10.1016/j.bbrc.2004.07.179) [Medline](#)
35. G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, E. S. Lander, Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming. *bioRxiv* 191056 [Preprint]. 27 September 2017. <https://doi.org/10.1101/191056>.

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 grants R33CA212697-01 and 1R01HL14102-01, the Harvard Stem Cell Institute Blood Program Pilot Grant DP-0174-18-00, and the Chan-Zuckerberg Initiative grant 2018-182714. A.R-F was supported by the Merck Fellowship of the Life Sciences Research Foundation, a 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 a NIH award K99HL146983. F.D.C was supported by NIH grants HL128850-01A1 and P01HL13147. F.D.C. is a scholar of the Howard Hughes Medical Institute and the Leukemia and 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: GSE140802. Further data are available at github.com/AllonKleinLab/paper-data.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.aaw3381/DC1

Materials and Methods

Figs. S1 to S12

Tables S1 to S8

References

10 December 2018; accepted 14 January 2020

Published online 23 January 2020

10.1126/science.aaw3381

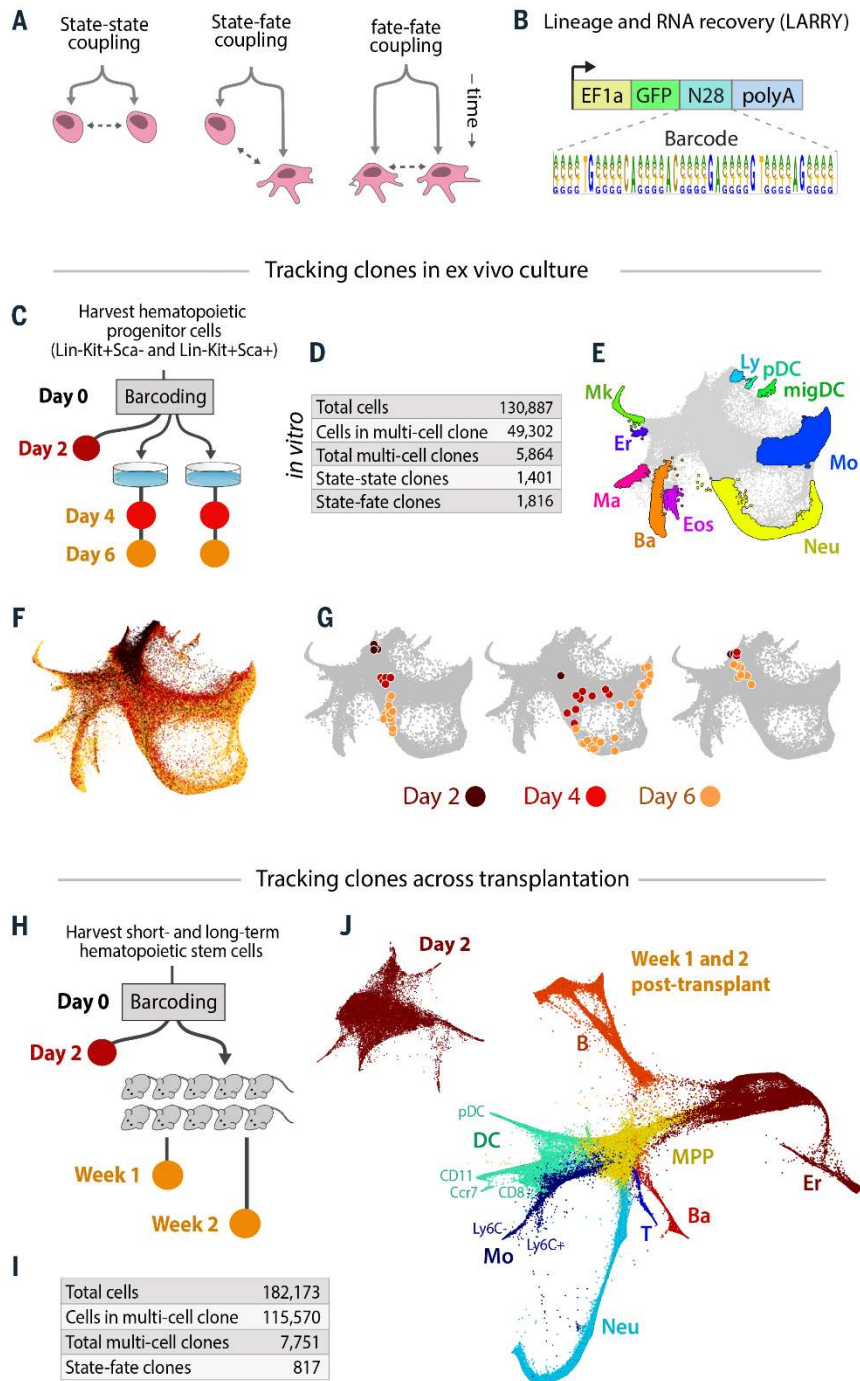


Fig. 1. Tracking clones over hematopoietic differentiation. (A) Experimental designs for tracking differentiation dynamics by analysis of sister cells. (B) The LARRY lentiviral construct delivers an expressed, heritable barcode that is detectable using scSeq. (C) Experiment tracking hematopoietic progenitor clones over time in primary culture. Colored circles indicate samples collected for scSeq. (D) Numbers of cells and clones sampled. (E) Annotated SPRING plot of transcriptomes from all time points [Ly=lymphoid precursor, Mk=megakaryocyte, Er=erythrocyte, Ma=mast cell, Ba=basophil, Eos=eosinophil, Neu=neutrophil, Mo=monocyte, DC=dendritic cell, migDC=migratory (ccr7+) DC, pDC=plasmacytoid DC]. (F) SPRING plot colored by 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) Experiment tracking clones after transplantation into 10 mice. Colored circles as in (C). (I) Numbers of cells and clones sampled. (J) scSeq data prior to transplantation (top-left) and post-transplantation (bottom-right), plotted as in (E) (T=T cell, B=B cell, NK=NK cell, MPP=multipotent progenitor).

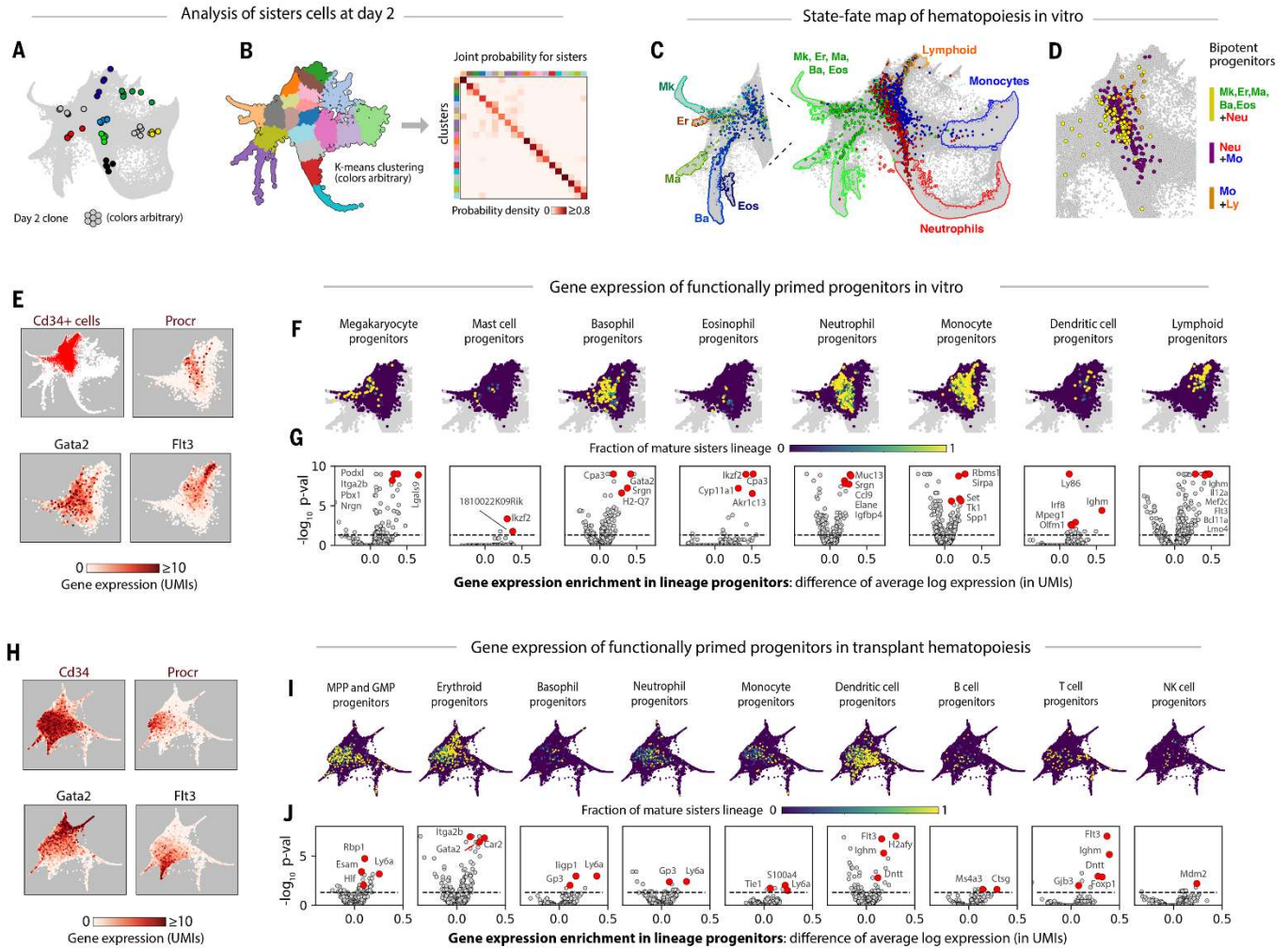


Figure 2. Linking state to fate in early hematopoiesis. (A and B) Sister cells at day 2 are transcriptionally similar as seen (A) by example (each color shows one clone), and (B) by the probability of sister cells occupying the same, or neighboring, transcriptional clusters. (C) Day 2 cells (colored 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 guides selection of early progenitors for further analysis. (F) Early progenitors colored by the fraction of sisters in each fate at days 4-6 in culture. (G) Volcano plots identify genes enriched among early progenitors for each lineage. Labeled genes shown red. (H) Heterogeneity among purified LSK cells after 2 days in culture prior to transplantation into mice. (I and J) Detection of early progenitor gene expression associated with future fates post-transplantation, repeating analyses from (E) to (G). In (E), (F), (I), and (J), points with non-zero value are plotted on top.

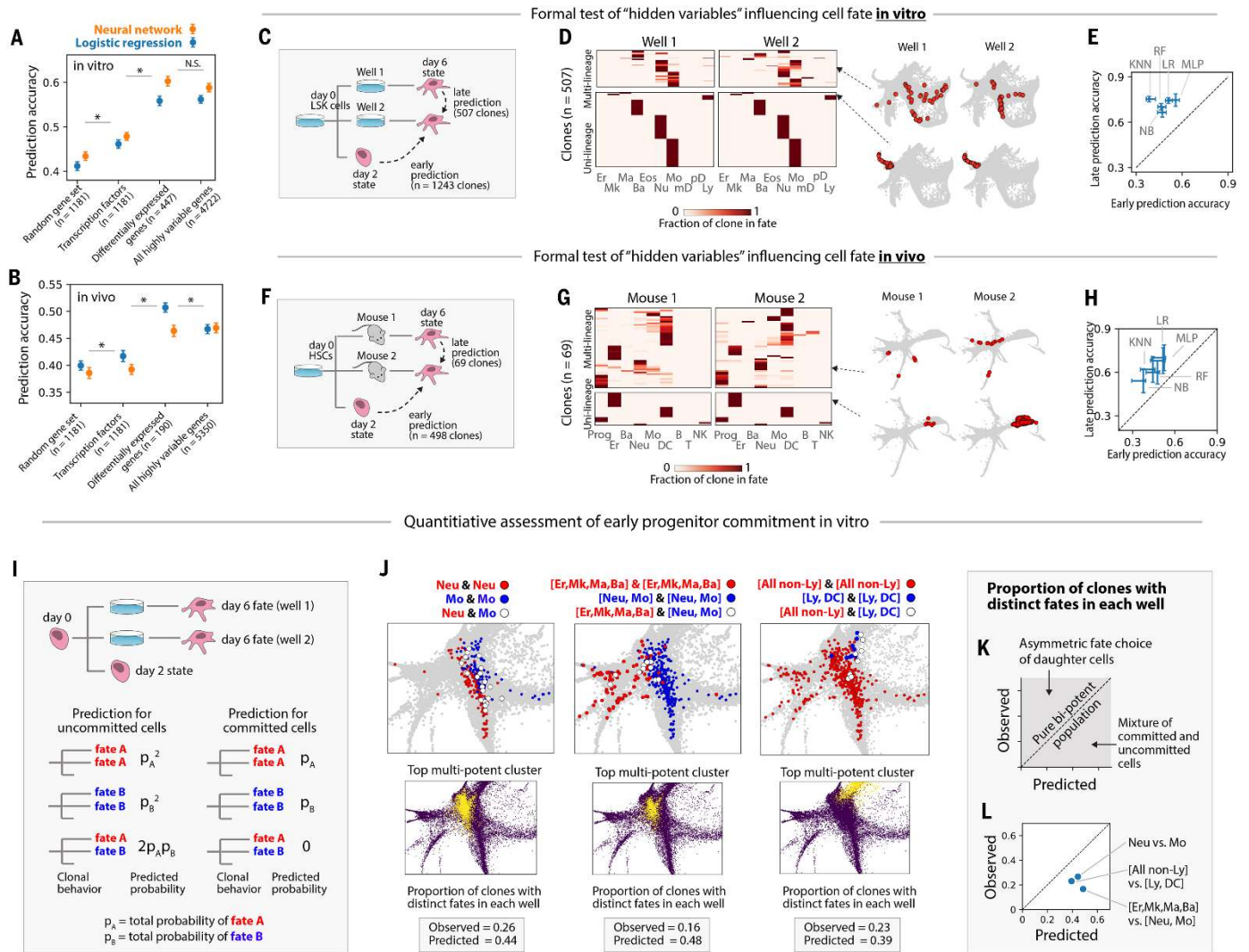


Fig. 3. Stochasticity and hidden variables from scSeq data. (A and B) Machine learning partially predicts clonal fate from the transcriptional state of early progenitors *in vitro* and *in vivo*. (Accuracy = fraction correct assignments). Asterisk (*) indicates statistical significance ($p < 10^{-4}$), 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, as compared to the early (day 2) state of a sister. (D and G) Clonal fate distributions for sisters 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 [naïve Bayes (NB), k-nearest neighbor (KNN), random forest (RF), multilayer perceptron (MLP)]. Error bars: standard deviation across 100 partitions of the data into training and testing sets. (I) A 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 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 vs. 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 pre-commitment. For all fate choices studied, fate priming or pre-commitment is inferred.

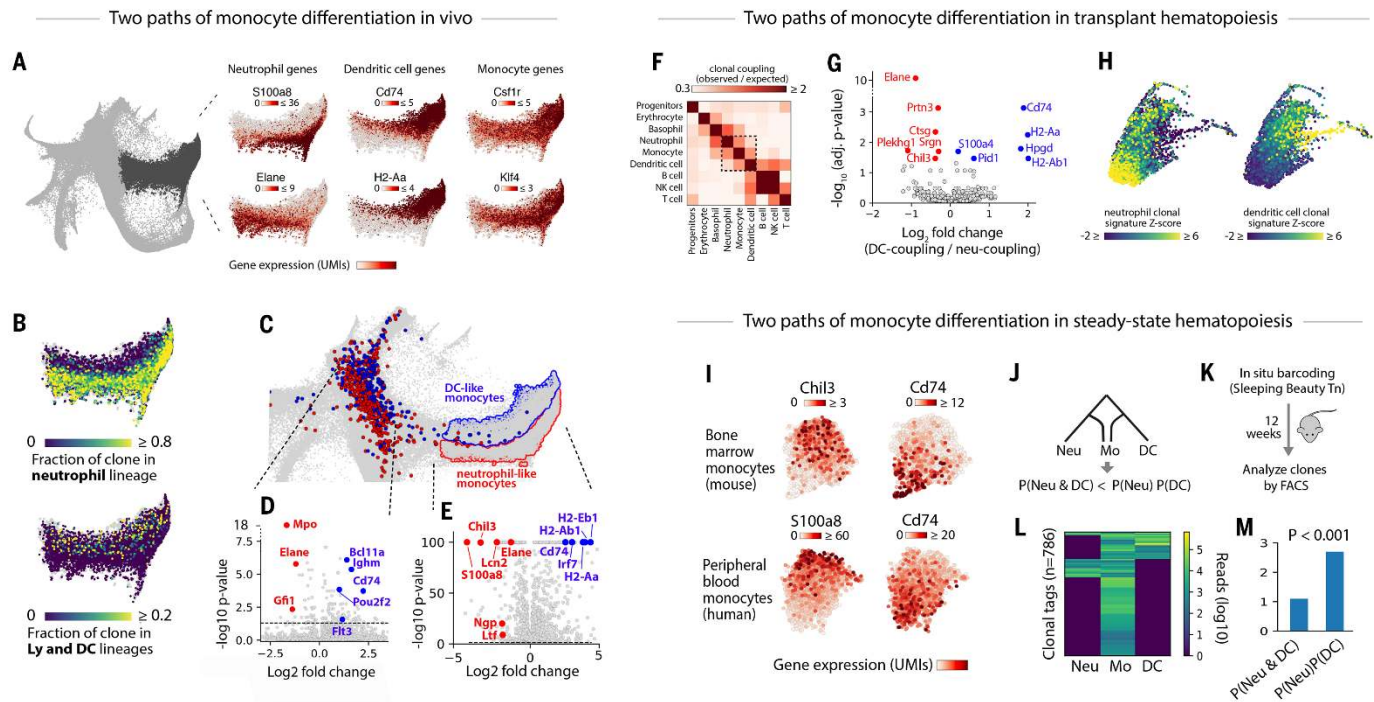
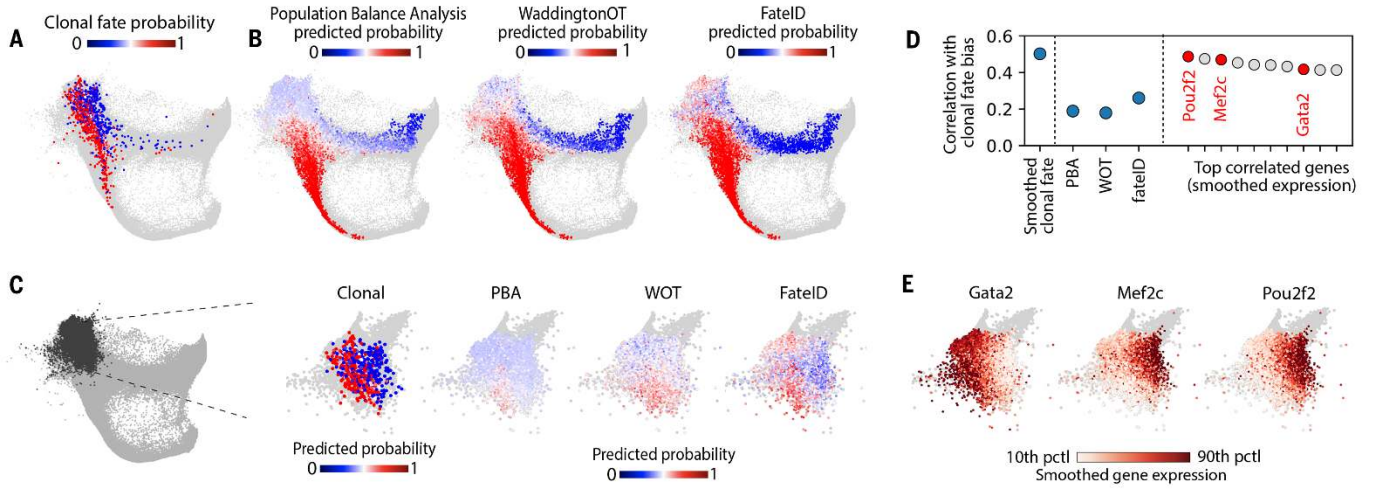
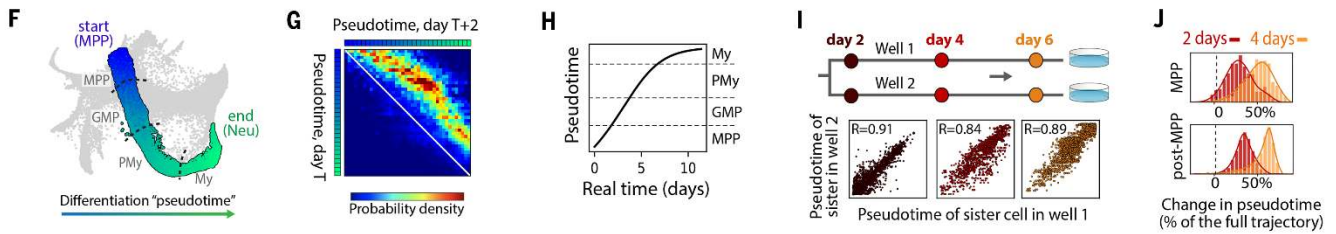


Fig. 4. Multiple paths of monocyte differentiation. (A) Differentiating monocytes show opposing expression of neutrophil and DC markers. Raw expression values are plotted with points ordered by expression level. (B) Monocytes segregate by proportions of neutrophil and DC sisters. Only monocytes for which 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 whose sisters differentiate into neutrophil-like or DC-like monocytes occupy distinct transcriptional states. Plot as in Fig. 2c. (D and E) Volcano plots identifying differentially-expressed genes between (D) the progenitors of, and (E) mature DC-like and neutrophil-like monocytes. (F) Barcodes overlap between cell-types indicates monocyte-DC and monocyte-neutrophil coupling one week post-transplantation. (G) Genes differentially expressed between monocytes related to neutrophils or to DCs after transplantation. (H) Signature scores (average of Z-scored expression) shown on a SPRING plot of post-transplantation monocytes. Points are ordered by expression level. (I) A DC-to-neutrophil axis of gene expression persists in mature monocytes, as seen by SPRING plots of scSeq data from monocytes in mouse bone marrow (top) and human blood (bottom). (J to M) Clonal analysis of monocyte differentiation in unperturbed hematopoiesis. (J) Under a model of two different monocyte 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) The number of cells in each type detected per clone (rows). (M) Observed vs. independent expectation for Mo-Neu-DC clones is consistent with two monocyte ontogenies.

Task 1: Cell fate prediction



Task 2: Temporal ordering (pseudotime)



Task 3: Lineage hierarchy

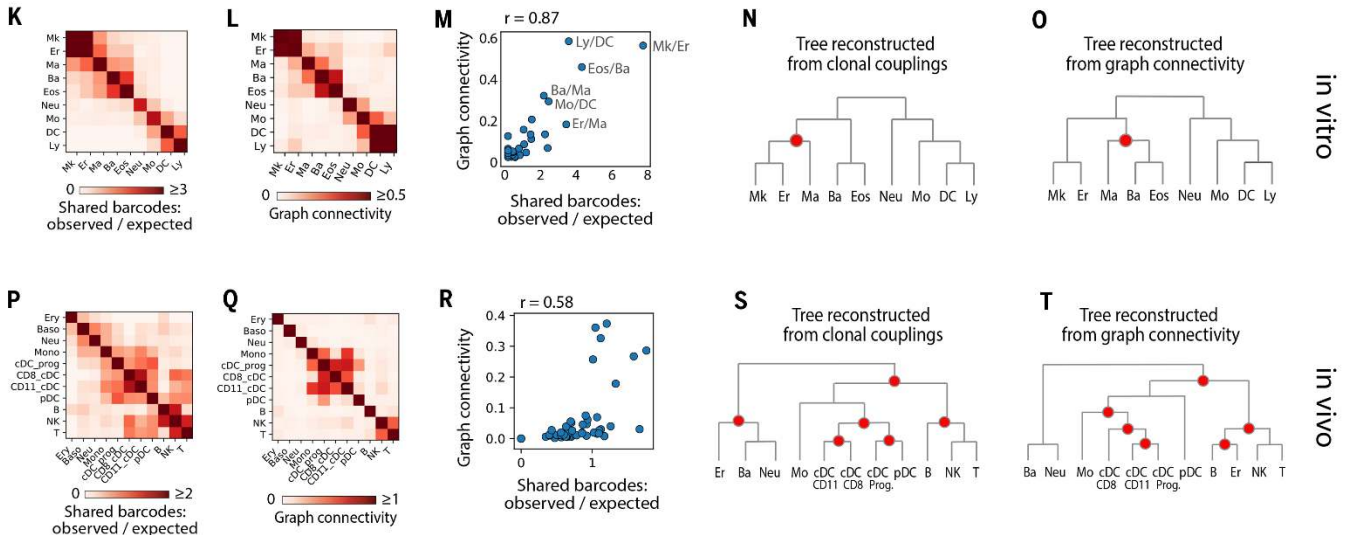


Fig. 5. A benchmark for dynamic inference from scSeq data. (A) SPRING plot of neutrophil/monocyte differentiation, with progenitors (day 2) colored by the ratio of neutrophil vs. monocyte fate of their sisters (days 4-6). (B) Algorithmic predictions of neutrophil vs. monocyte fate from transcription alone fail to recognize the early fate boundary revealed by clonal tracking. (C) Expanded view of early progenitors (thresholded by CD34 expression); plots as in (A) and (B). (D) Pearson correlation between future clonal fate outcomes of early progenitors and (I) smoothed fate probabilities of held-out clonal data, (II) output of algorithmic predictions, (III) expression of top 10 most-correlated genes (red = transcription factors). Held-out data sets 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 neutrophil 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 show greater variability in MPPs compared to later stages (red=day 2 to 4; orange=day 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 k-nearest-neighbor 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 based on the clonal or state distances. Red dots indicate the sole discrepancy between the hierarchies. (P to T) As (K to O) repeated for cells post-transplantation, showing increased discrepancies between clonal and state-based hierarchies.