Single-Cell Gene Expression Profiling and Cell State Dynamics: Collecting Data, Correlating Data Points and Connecting the Dots

Carsten Marr[1,$], Joseph X. Zhou[2,$], Sui Huang[2,*]

[1]*Helmholtz Zentrum München - German Research Center for Environmental Health, Institute of Computational Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany*

[2]*Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109, USA*

[$]*These authors contributed equally*

[*]*Corresponding author: Sui Huang (sui.huang@systemsbiology.org)*

Abstract

Single-cell analyses of transcript and protein expression profiles – more precisely, single-cell resolution analysis of molecular profiles of cell populations – have now entered center stage with the wide application of single-cell qPCR, single-cell RNA-Seq and CytOF. These high-dimensional population snapshots techniques are complemented by low-dimensional time-resolved microscopy-based monitoring methods of individual cells. Both fronts of advance have exposed a rich heterogeneity of cell states within uniform cell populations in many biological contexts, producing a new kind of data that has stimulated a series of computational analysis methods for data visualization, dimensionality reduction, and "cluster" (subpopulation) identification. The next step is to go beyond collecting data and correlating data points with computational analyses: to connect the dots, that is, to understand what actually underlies the identified data patterns. This entails interpreting the "clouds of points", each representing a cell in state space, and their structure as manifestation of the regulation by the molecular network. This control of cell state dynamics can be formalized as a quasi-potential landscape, as first proposed by Waddington. We summarize not only key methods of data acquisition and computational analysis but also explain the principles that link the single-cell resolution measurements to dynamical systems theory.

Introduction

A phenotype switch of a cell, or more formally, a cell state transition, is an elementary event in metazoan development. The associated phenotype change, e.g. cell differentiation, cell growth termination or artificial cell reprogramming, has traditionally been explained by molecular signaling pathways. This understanding has been extended to exhaustively characterizing *cell states* - defined by molecular profiles, such as transcriptomes or proteomes. However, the characterization of static molecular profiles cannot explain essential properties of the cell state *dynamics*, such as discreteness of states, stability of states, binary nature of cell transitions and the directionality of cell development. These properties emerge from nonlinear dynamics of the molecular regulatory networks involved, such as the gene regulatory network that governs cell state dynamics.

Only in the past decade has nonlinear dynamical systems theory entered the center stage in the study of cell state transitions with the renaissance of Waddington's epigenetic landscape [1] as a conceptual aid in stem cell and developmental biology. Waddington proposed a "landscape" to explain that cells differentiate into discrete, robust cell states. In his view, a marble, representing a cell, rolls down a hilly landscape towards a number of valleys and must eventually settle in one of them - each representing a particular cell type. This landscape is, as we will see,more than a metaphor but has a mathematical basis.

With the arrival of single-cell technologies we can in principle uncover the topography of the landscape by profiling individual cells in as many positions as possible. Technologies for monitoring single-cell states can be divided into two complementary types: measurement (i) of a large number of variable of a cell state (e.g. abundance of transcripts/proteins) as “snapshot” at a given time point in a large number of cells or (ii) of just a handful of variables continuously over time in the same cell and its descendants. The former destroys the cells during measurement, the latter keeps cells alive, allowing for longitudinal monitoring and providing information unique to biological systems, such as dependencies between mother and daughter cells. While the analysis of single developing cells has a over 100 year old history (see [2] for a recent review) a, what is new is the massively parallel nature and the high-dimensionality: a large number of cells can be analyzed simultaneously for a large number of cellular variables. Thus, novel technologies are less about “single-cell”, but rather allow analyses of entire cell populations with single-cell *resolution*.

So far most analyses of single-cell profiles and longitudinal observations are agnostic of the formalism of dynamical systems that underlies the intuitive picture of the landscape. Although sometime Waddington’s landscape is invoked, current approaches almost exclusively focus on descriptive computational analysis for data visualization, dimension reduction, or statistical pattern identification (see [3,4] for reviews). But now the time is ripe to move beyond collecting the data and correlating the data points, and to connect the dots: We need to unite the formal theory of dynamical systems that explains the uncovered patterns in single-cell data. In the following, we review recent technological developments (collecting the data) and the computational tools as a first level data organization (correlating the data points) before summarizing the interpretation of data in the light of formal concepts of dynamical systems (connecting the dots).

Collecting the Data: Snapshot sampling of single cells

Single-cell techniques, if applied to a sufficiently large number of cells, provide distributions of measured variables for the entire population. Such distributions offer unprecedented wealth of information about the dynamics of cell states, far beyond cell-cell variability and higher statistical moments.

To appreciate this, one needs first to accept the biological fact that variable distribution even in an isogenic uniform population of cells of the nominally same “cell type” do not just reflect inconsequential (thermal) fluctuations in gene expression, let alone technical (measurement) noise, but also a biologically significant diversity of cellular states with functional consequences. For instance, fluorescent activated cell sorting (FACS) and analysis have revealed distinct subpopulation dynamics among mouse embryonic stem cells (mESCs) with respect to expression of the pluripotency factor Nanog [5]. It is also evident that individual cell states are not static but dynamic, exposed by the slow noise-driven re-establishment of a heterogeneous marker distribution in hematopoietic cells from a sorted subpopulation [6]. The main limitation for FACS is that the number of proteins that can be simultaneously analyzed barely exceeds a dozen due to overlap of optical emission spectra. An advance is single cell mass cytometry (CyTOF), where antibodies are tagged with heavy metals and measured with mass spectrometry [7–9]. The sharper discrimination of labels allows for up to 50 proteins to be measured simultaneously in each cell.

Single-cell technologies for measuring transcripts also progressed tremendously in recent years. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) technology on nanoliter-scale has been applied to nearly 4,000 cells in different stages of early blood development [10]. However, the number of different mRNAs that can be analyzed in one cell is limited in PCR-based approaches. Profiling whole transcriptomes can be achieved with single-cell mRNA sequencing (RNASeq), where the crucial step is unbiased amplification of cDNA before sequencing (see [11–13] for reviews). Use of unique molecular identifiers (UMIs) that bar-code each molecule, not just each transcript species allows a robust quantification by intercepting amplification bias [14].

Combining several new methods (barcoding for multiplexing cells to run in the same sequencing reaction, use of microfluidics and droplets for initial reactions of individual cells), cell throughput has been pushed upwards (see Figure 1) recently. Using DropSeq, 39 subpopulations of mouse retinal cells have been identified in 40,000 cells [15] and cell population heterogeneity in nearly 6,000 mESCs has been profiled [16]. A drawback in RNASeq as compared to qRT-PCR is the reduced sensitivity such that only the ~10% most abundant transcripts can be quantitated [17]. Careful experimental design, balancing the tradeoff between number of cells and transcripts sequenced and sequencing depth [17] and appropriate computational post-processing, e.g. to correct for cell-cycle induced heterogeneities [18], are thus crucial for single cell transcriptomics.

Gene expression manifests cell states only at one level. The chromatin state of single cells can be determined for individual cells with single-cell bisulfite sequencing for DNA methylation [19], the Hi-C method for chromosome conformation [20] transposase-accessible chromatin assay (ATAC-seq) [21,22], and immunoprecipitation followed by sequencing (ChIP-seq) for histone methylation [23]. Also the spatial component is now accessible with recent extensions to fluorescent in situ hybridization [24,25].

Correlating the data points: Computational data analysis of Snapshots

High-throughput (many cells) and high-dimensional (many variables) single-cell measurement provides a wealth of information on molecular profiles in populations. The output for a nominal biological sample point (e.g. a time after drug treatment) is not a vector of m components (measured variables like mRNA species) as in transitional population/tissue omics, but a matrix of [$m \times n$] since we now measure n cells. Mathematically, each cell can be positioned in a m dimensional space, where the axes are the measured variables. Using this notation, which is also the basis for a dynamical systems analysis discussed later, the first generation of computational tools has been developed to handle this new type of data: to reduce the m-dimensional space by mapping individual cells onto an interpretable lower (two- or three-) dimensional space with minimal loss of information, to identify patterns, such as an (pseudo)temporal order of cell states or static clusters, or just to visualize the data. Many approaches are in principle identical to those used in conventional transcriptome analysis, except that instead of a sample vector $\boldsymbol{x}$ at sample point T of tissues or bulk cell populations, we now have n cells, each with their states at the sample points T.

Pearson's principal component analysis (PCA) identifies genes that vary most within the profiled population of cells and linearly projects the high-dimensional data into a lower dimensional space. It has been used for instance, to classify mouse sensory neurons into novel cellular subtypes [26]. The incorporation of censoring of expression values due to non-detected transcripts has been achieved by a probabilistic version of PCA [27] and an extension of the factor analysis framework [28]. Nonlinear methods are often better suited to identify the low-dimensional manifolds in which RNA expression is confined. One popular method is t-SNE - a variant of stochastic neighbor embedding using the Student-t distribution to calculate pointwise similarity [29]. Recently, it has been used to cluster retinal cells [15] and embryonic stem cell populations [16] from droplet-based RNASeq. t-SNE preserves local distances and thus ensures that neighboring data points in the original data space are still nearby in the low-dimensional embedding. However points distant in the data space could be rather close-by in the embedding. Alternative dimension reduction have been applied, based on Gaussian processes [30] and diffusion maps [31,32], and preserve global state space distances between cells. A comparison of several dimension-reduction algorithms for single-cell analysis is provided in [31].

Several approaches have been developed to identify clusters within the high-dimensional data set, a first step towards discovering new cell subpopulations of biological significance. Spectral clustering [33] and density-based cell population identification [34] for the analysis of FACS data has been

proposed. For CyTOF data, subpopulations have been identified with a regularized regression-based method [35] and a graph-based method with community detection to maximize “modularity” [36]. Cell hierarchies have been estimated based on minimum spanning trees [37], and a divisive bi-clustering method [38] infers classes of molecularly distinct cells in the mouse brain. Of obvious biological interest is also to identify rare cell types outside of abundant clusters. Gruen et al. [39] devised an algorithm to do so and predicted and validated a rare intestinal cell type.

A first hint of awareness of a dynamical process underlying the observed patterns is offered by methods that seek to quantify interrelatedness between cells by viewing them as snapshots of a temporal succession of states despite being measured at the same time. Such “smearing out” of a population is plausible given the stochastic asynchrony of biological processes between cells. The ‘monocle’ package [40] allows the inference of a pseudotemporal order via independent component analysis (ICA) and reconstruction of a mean spanning tree based path through the low dimensional embedding. An alternative algorithm called Wanderlust [41] reconstructs a developmental trajectory based on nearest neighbor graphs in the high-dimensional measurement space. A statistical analysis was used to infer oscillatory genes from a single-cell RNASeq snapshot, where unsynchronized cells were mixed [42].

Longitudinal sampling

All methods above destroy the observed samples, or lose the identity of individual cells between two measurement points. Some problems [43] require the monitoring of the molecular state of the very same cell at distinct time points. This is achieved by traditional video microscopy or, if cytotoxicity is involved, and the process of interest is slow, by time-lapse microscopy. Successful implementation requires (i) choice of appropriate markers, (ii) conditions that keep cells in physiological conditions under the microscope and (iii) suitable methods for tracking of individual cells (see [2,44] for reviews).

In mESCs, the dynamics of the heterogeneously expressed pluripotency factor Nanog has been scrutinized in this manner, demonstrating the switching between Nanog-high and Nanog-low cells. Such longitudinal studies quantified gene expression fluctuations across cell cycles, subpopulations, and conditions using reporter systems [45–49] or fusion proteins [50]. To study fundamental aspects of gene expression, mRNA levels have be measured using the MS2 system [46], where a specific target sequence is incorporated in the non-coding portion of the RNA of interest, forming a RNA stem loop that is bound by the constitutively expressed viral MS2 protein fused with a fluorescent protein [51]. Naturally, these live cell imaging methods measure low-dimensional dynamics: they monitor a relatively small number (typically up to two) of state variables, but in turn they provide a wealth of information on the temporal structure of gene expression.

Connecting the dots: Analysis informed by Dynamical Systems Theory

The patterns discovered by descriptive computational analyses, and how they change as cells switch phenotypic state, must obviously be driven by a force and guided by constraints. A given single cell state, for instance, can be stable or unstable, fated toward a particular state or capable of choosing between multiple fates [52]. The governing principles of these dynamical properties can be comprehended in a formal manner by considering the regulatory network that controls gene expression. With that we move from the epistemology of phenomenological analysis to that of a formal explanation framework anchored in a set of first principles.

It is in this sense that Waddington‘s landscape enters the interpretation of the single-cell data. The landscape manifests the constraints on cell state changes (dx/dt) that emanate from the nonlinear dynamics of the underlying regulatory system (dx/dt = F(x)) [1,53–55], which in theoretical studies is often embodied by a gene regulatory network (for network inference based on snapshot data see [10,56]). This is because the principle of how molecular interactions coordinate gene expression to

define cell states are conceptually understood, and cell states are well-represented in the readily measurable transcriptome. The regulatory interactions between gene loci generate a driving force on the cell: For instance, if gene A encodes an inhibitor of gene B, then as A increases, B will decrease. Such coordinated change of expression across all loci in the genome takes place until all interactions are "satisfied" (dx/dt = F(x) = 0) and the driving force vanishes. This state corresponds to a stable attractor – the point at the bottom of a potential well. Thus, in principle, the entire behavioral repertoire of a cell system is encoded in the genome and uniquely maps into a particular landscape [54]. Internal fluctuations of gene expression push the cell away from the attractors – "against the uphill slope". But if the restoring force is not exceeded, the cell will be pushed back to the attractor and on average all cells of a population wiggle around an attractor state. This confluence of deterministic and stochastic dynamics is what gives rise to "clusters" in state space, which account for the non-genetic cell-cell variability (heterogeneity) within a population of cells from the nominally same type. The spread of the clusters around the attractor state is a measure of heterogeneity [57]. "Nonlinearity" of a system implies that for elementary mathematical reasons, the rate equations that describe the dynamics of the network can produce multiple solutions – that is, multiple attractor states [58].

In the modern version of Waddington's landscape [59] we are interested in the "relative" stability of each attractor – the relative "depth" of the valleys. In this perspective, different cell types have distinct quasi-potentials. This is important because a phenotype change induced by external perturbation can be imagined as an equivalent of "catalysis": the barrier (hill) between attractors is lowered because of changes in the regulatory interactions conferred by signal transduction. This allows cells to swarm out of the original, now flattening attractor and reach new nearby attractor states [60] (see Figure 3). Thus even from the qualitative landscape image the "relative depth" of an attractor, akin to (but fundamentally distinct from) chemical potential wells, governs the direction of likely transitions. The landscape slope embodies the driving force of cell differentiation and arrow of time of development [61]. A landscape in which the quasi-potential U (elevation) reflects the probability of transitions between attractors along a "least effort path" can in theory be numerically computed exactly; but this would require knowledge of the system specifications $F(x)$ from the governing rate equations $dx/dt=F(x)$ of the dynamical system, that is, the architecture of the network and reaction modalities of every regulatory interaction. Since such detailed knowledge is not available and constructing $U(x)$ may be computationally extremely expensive even if we had the information, only partial landscapes can be derived from models of known gene-gene interactions that form small circuits [62].

However, single-cell technology and the measurement of high-dimensional states of many cells now provide a way to determine the *relative occupancy* probabilities of attractor states (density of clusters in state space) and *attractor transition rates* (at which a cell moves from one cluster to another). From these two measurements, we can phenomenologically obtain the landscape shape, such as relative sizes and depths of attractors, and height of barriers between them, directly from single cell states without knowledge of the specification of the dynamical system. The general idea is that the stochasticity of individual cells turns a cell population into a statistical ensemble that "reads out" the constrained state space as imposed by the gene regulatory network. For instance from the cell density distribution in state space and at steady-state, we can define attractors. The transition rates between attractors can be revealed by sorting cells from one cluster and observing transitions to reconstitute another [6]. According to these transition rates one can estimate their "relative stability" based on the theory of quasi-potential energies. A widely used intuitive approximation of the depth of attractor is $U \sim ln(P_{ss})$, where P_{ss} is the measured density of states [63,64]. Note, however that a difference in this apparent potential is that this is not the source of the force that drives the state change: given the rate of state change and the quasi-potential $U \sim ln(P_{ss})$, the driving force is not $F = -grad\, U$ [55]. Exact experimental determination of the probabilities of transitions between attractors would require longitudinal monitoring, in all relevant state dimensions, of cells undergoing such transitions – which is currently beyond reach. However, cell transition rates in lower dimension can be measured [50] and have been modeled with a phenomenological bifurcation model [65].

Finally, a theory-based interpretation of snapshot data is based on the concept of ergodicity. In analogy to the ergodic hypothesis in statistical physics, a static picture of an ensemble of individual cells can inform us about the behavior of an individual over time – if the state changes of individuals are fast enough relative to the time of observation. The ergodic rate analysis (ERA) was used to estimate the rates of cell growth based on snapshots of a cell population [66], and population dynamics was deconvolved by tracking individual cells after cancer drug treatment [67].

Conclusions

To understand the cell population substructures manifesting the constraints imposed by the underlying dynamical system which can be intuitively and formally depicted as a quasi-potential landscape, we need to go beyond descriptive computational data analysis and *ad hoc* interpretation and enter the still little charted terrain of theory-based interpretation. Here the most natural framework for understanding why in the first place the patterns in the data arise, is the theory of nonlinear stochastic dynamical systems [68]. In the near future we will see progress at all three fronts discussed in this article: In collecting data, the costs for profiling individual cells will drop drastically. For instance, DropSeq for transcript measurement will make RNAseq with ten thousands of cells affordable, which is critical for statistically robust evaluation of population substructure. At the front of correlating data points we expect to see a consolidation. First-generation computational tools have served their purpose in introducing the intuition of single-cell resolution analysis of high-dimensional cell-states, but lack a deeper understanding of the underlying regulatory system. At the front of connecting the dots, theory-based analysis will benefit from progress and sinking costs in data collection which will permit the design of more complex experimental schemes, with denser snapshots in order to test the theory. However, theoretical concepts, such as the quasi-potential landscape must be further developed and linked to data, and the abstract ideas need to be disseminated to a larger community of bioinformaticians.

Acknowledgements

We thank Chris McGinnis (Seattle), and Michael Strasser, Alex Wolf, and Thomas Blasi (Munich) for comments on the manuscript, Philipp Angerer (Munich) for technical support, and the German Academic Exchange Service DAAD and Bavarian Research Alliance BayFOR for exchange funding. Research reported in this publication was supported by the National Institutes of Health under award number R01GM987654. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Figures

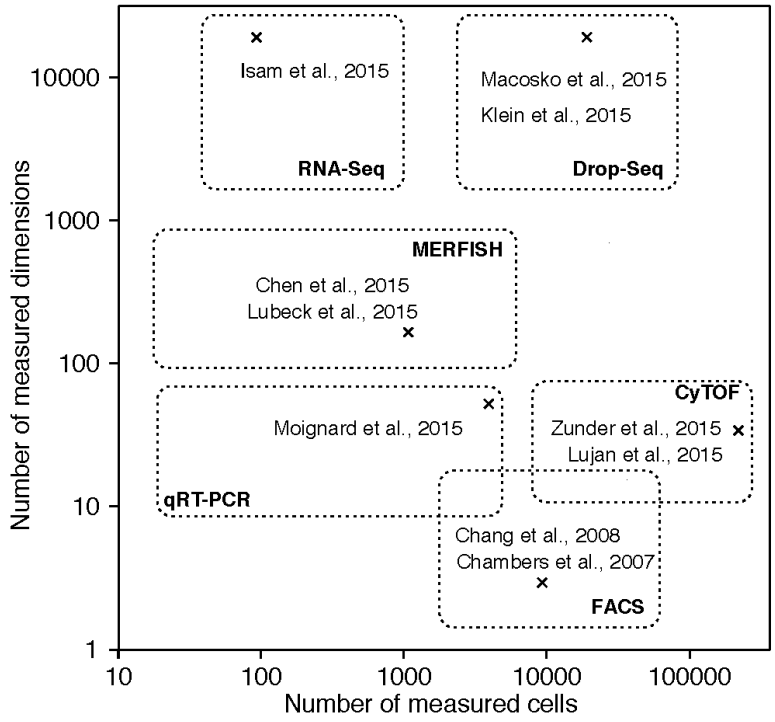

Figure 1: Recent applications of single-cell snapshot technologies sample up to hundreds of thousands of single cells while measuring multiple variables, e.g. genome-wide mRNA expression.

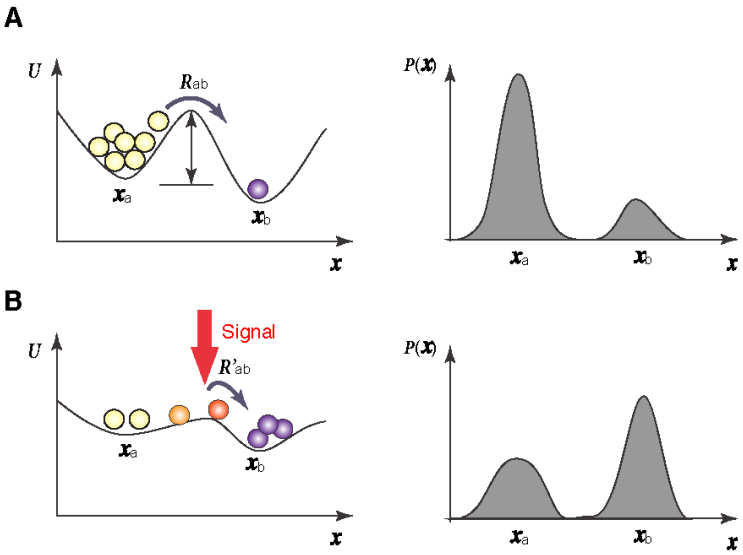

*Figure 2: A cell state transition is driven by the quasi-potential (landscape) change induced by external signals. In (**A**)* a bistable landscape exists with cells residing mainly in state x_{a} due to a high barrier to state x_{b}. *In (**B**)* cells transit to x_{b}, *caused by the external signals that flatten the barrier.*

References

1. Waddington CH: **Principles of embryology**. *London: George Allen & Unwin Ltd.* 1956, [no volume].

2. Coutu DL, Schroeder T: **Probing cellular processes by long-term live imaging--historic problems and current solutions**. *J. Cell Sci.* 2013, **126**:3805–3815.

3. De Vargas Roditi L, Claassen M: **Computational and experimental single cell biology techniques for the definition of cell type heterogeneity, interplay and intracellular dynamics**. *Curr. Opin. Biotechnol.* 2015, **34**:9–15.

4. Trapnell C: **Defining cell types and states with single-cell genomics**. *Genome Res.* 2015, **25**:1491–1498.

5. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A: **Nanog safeguards pluripotency and mediates germline development**. *Nature* 2007, **450**:1230–1234.

6. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S: **Transcriptome-wide noise controls lineage choice in mammalian progenitor cells**. *Nature* 2008, **453**:544–547.

The first paper to show that non-genetic gene expression heterogeneity within a clonal population creates metastable states.

7. Krishnaswamy S, Spitzer MH, Mingueneau M, Bendall SC, Litvin O, Stone E, Pe'er D, Nolan GP: **Systems biology. Conditional density-based analysis of T cell signaling in single-cell data**. *Science* 2014, **346**:1250689.

8. Zunder ER, Lujan E, Goltsev Y, Wernig M, Nolan GP: **A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry**. *Cell Stem Cell* 2015, **16**:323–337.

Molecular profiling of reprogramming dynamics of 36 markers in 250,000 mouse embryonic stem cells using mass cytometry.

9. Lujan E, Zunder ER, Ng YH, Goronzy IN, Nolan GP, Wernig M: **Early reprogramming regulators identified by prospective isolation and mass cytometry**. *Nature* 2015, **521**:352–356.

10. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, Buettner F, Macaulay IC, Jawaid W, Diamanti E, et al.: **Decoding the regulatory network of early blood development from single-cell gene expression measurements**. *Nat. Biotechnol.* 2015, **33**:269–276.

Gene expression profiling of nearly 4,000 mouse embryonic cells with blood forming potential using qRT-PCR. Computational analysis using diffustion maps and network reconstruction.

11. Junker JP, van Oudenaarden A: **Every cell is special: genome-wide studies add a new dimension to single-cell biology**. *Cell* 2014, **157**:8–11.

12. Saliba A-E, Westermann AJ, Gorski SA, Vogel J: **Single-cell RNA-seq: advances and future challenges**. *Nucleic Acids Res.* 2014, **42**:8845–8860.

13. Sandberg R: **Entering the era of single-cell transcriptomics in biology and medicine**. *Nat. Methods* 2014, **11**:22–24.

14. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S: **Quantitative single-cell RNA-seq with unique molecular identifiers**. *Nat. Methods* 2014, **11**:163–166.

Barcodes for individual molecules shown to allow for alleviating amplification noise and the assessment of transcript abundances.

15. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al.: **Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets**. *Cell* 2015, **161**:1202–1214.

Introduction of Drop-seq, a droplet-microfluidic that allows handling and simultaneous sequencing of thousands of cells, applied to the identification of distinct cell populations in over 44,000 mouse retinal cells.

16. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW: **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells**. *Cell* 2015, **161**:1187–1201.

Introduction of InDrop, similar to Drop-seq, applied to the characterization of mouse embryonic stem cell subpopulations.

17. Stegle O, Teichmann SA, Marioni JC: **Computational and analytical challenges in single-cell transcriptomics**. *Nat. Rev. Genet.* 2015, **16**:133–145.

18. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells**. *Nat. Biotechnol.* 2015, **33**:155–160.

Computational approach to account for confounding factors in single cell gene expression data, applied to cell cycle induced variations.

19. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G: **Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity**. *Nat. Methods* 2014, **11**:817–820.

20. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure**. *Nature* 2013, **502**:59–64.

21. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ: **Single-cell chromatin accessibility reveals principles of regulatory variation**. *Nature* 2015, **523**:486–490.

22. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J: **Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing**. *Science* 2015, **348**:910–914.

23. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, Bernstein BE: **Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state**. *Nat. Biotechnol.* 2015, doi:10.1038/nbt.3383.

24. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L: **Single-cell in situ RNA profiling by sequential hybridization**. *Nat. Methods* 2014, **11**:360–361.

25. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X: **RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells**. *Science* 2015, **348**:aaa6090.

26. Usoskin D, Furlan A, Islam S, Abdo H, Lönnerberg P, Lou D, Hjerling-Leffler J, Haeggström J,

Kharchenko O, Kharchenko PV, et al.: **Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing**. *Nat. Neurosci.* 2015, **18**:145–153.

27. Buettner F, Moignard V, Göttgens B, Theis FJ: **Probabilistic PCA of censored data: accounting for uncertainties in the visualization of high-throughput single-cell qPCR data**. *Bioinformatics* 2014, **30**:1867–1875.

28. Pierson E, Yau C: **ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis**. *Genome Biol.* 2015, **16**:241.

29. Van der Maaten L, Hinton G: **Visualizing data using t-SNE**. *J. Mach. Learn. Res.* 2008, **9**:85.

30. Buettner F, Theis FJ: **A novel approach for resolving differences in single-cell gene expression patterns from zygote to blastocyst**. *Bioinformatics* 2012, **28**:i626–i632.

31. Haghverdi L, Buettner F, Theis FJ: **Diffusion maps for high-dimensional single-cell analysis of differentiation data**. *Bioinformatics* 2015, **31**:2989–2998.

Application of the diffusion map concept to transcriptomics data of differentiation cells, including an overview of existing dimension reduction algorithms.

32. Angerer P, Haghverdi L, Büttner M, Theis F, Marr C, Buettner F: **destiny--diffusion maps for large-scale single-cell data in R**. *bioRxiv* 2015.

33. Zare H, Shooshtari P, Gupta A, Brinkman RR: **Data reduction for spectral clustering to analyze high throughput flow cytometry data**. *BMC Bioinformatics* 2010, **11**:403.

34. Malek M, Taghiyar MJ, Chong L, Finak G, Gottardo R, Brinkman RR: **flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification**. *Bioinformatics* 2015, **31**:606–607.

35. Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP: **Automated identification of stratifying signatures in cellular subpopulations**. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**:E2770–7.

36. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir E-AD, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER, et al.: **Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis**. *Cell* 2015, **162**:184–197.

37. Qiu P, Simonds EF, Bendall SC, Gibbs KD Jr, Bruggner RV, Linderman MD, Sachs K, Nolan GP, Plevritis SK: **Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE**. *Nat. Biotechnol.* 2011, **29**:886–891.

38. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, et al.: **Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq**. *Science* 2015, **347**:1138–1142.

39. Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A: **Single-cell messenger RNA sequencing reveals rare intestinal cell types**. *Nature* 2015, **525**:251–255.

40. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells**. *Nat. Biotechnol.* 2014, **32**:381–386.

Introduction of the Monocle method for pseudotemporal ordering of differentiating human muscle progenitor cells.

41. Bendall SC, Davis KL, Amir E-AD, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Pe'er D: **Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development**. *Cell* 2014, **157**:714–725.

Introduction of the Wanderlust algorithm that aligns single cells along a high-dimensional trajectory to reconstruct the developmental path of human white blood cell maturation.

42. Leng N, Chu L-F, Barry C, Li Y, Choi J, Li X, Jiang P, Stewart RM, Thomson JA, Kendziorski C: **Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments**. *Nat. Methods* 2015, doi:10.1038/nmeth.3549.

43. Pisco AO, Brock A, Zhou J, Moor A, Mojtahedi M, Jackson D, Huang S: **Non-Darwinian dynamics in therapy-induced cancer drug resistance**. *Nat. Commun.* 2013, **4**:2467.

44. Hoppe PS, Coutu DL, Schroeder T: **Single-cell technologies sharpen up mammalian stem cell research**. *Nat. Cell Biol.* 2014, **16**:919–927.

45. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, Cai L, Elowitz MB: **Dynamic heterogeneity and DNA methylation in embryonic stem cells**. *Mol. Cell* 2014, **55**:319–331.

Quantitative time-lapse microscopy of transcriptional Nanog reporter in mouse embryonic stem cell reveals stochastic switching between gene expression states.

46. Ochiai H, Sugawara T, Sakuma T, Yamamoto T: **Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells**. *Sci. Rep.* 2014, **4**:7125.

Time-lapse data used to infer the Nanog DNA state from mRNA and protein kinetics in mouse embryonic stem cells.

47. Abranches E, Guedes AMV, Moravec M, Maamar H, Svoboda P, Raj A, Henrique D: **Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency**. *Development* 2014, **141**:2770–2779.

48. Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser AJ, Li H, Zhang J, Pardee K, Gennert D, Trombetta JJ, et al.: **Deconstructing transcriptional heterogeneity in pluripotent stem cells**. *Nature* 2014, **516**:56–61.

Landscape of gene expression variability determined by intercolony variance in transcript numbers of several pluripotency and lineage regulators with fluorescence in situ hybridisation (FISH), after 3-4 days of clonal growth, in traditional *in-vitro* and perturbed conditions

49. Cannon D, Corrigan AM, Miermont A, McDonel P, Chubb JR: **Multiple cell and population-level interactions with mouse embryonic stem cell heterogeneity**. *Development* 2015, **142**:2840–2849.

50. Filipczyk A, Marr C, Hastreiter S, Feigelman J, Schwarzfischer M, Hoppe PS, Loeffler D, Kokkaliaris KD, Endele M, Schauberger B, et al.: **Network plasticity of pluripotency transcription factors in embryonic stem cells**. *Nat. Cell Biol.* 2015, doi:10.1038/ncb3237.

Analysis of cell state stability on the protein level, and transitions between them with thousands of single cells and millions of timepoints from quantitative time-lapse microscopy.

51. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM: **Localization of ASH1 mRNA particles in living yeast**. *Mol. Cell* 1998, **2**:437–445.

52. Huang S: **Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways**. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2011, **366**:2247–2259.

53. Kauffman SA: **Metabolic stability and epigenesis in randomly constructed genetic nets**. *J.*

Theor. Biol. 1969, **22**:437–467.

54. Huang S: **Cell Lineage Determination in State Space: A Systems View Brings Flexibility to Dogmatic Canonical Rules**. *PLoS Biol.* 2010, **8**:e1000380.

55. Zhou JX, Aliyu MDS, Aurell E, Huang S: **Quasi-potential landscape in complex multi-stable systems**. *J. R. Soc. Interface* 2012, **9**:3539–3553.

Detailed theoretical construction of quasi-potential landscape based on Wentzell large deviation theory and its difference from U ~ ln P_{ss}

56. Ocone A, Haghverdi L, Mueller NS, Theis FJ: **Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data**. *Bioinformatics* 2015, **31**:i89–96.

57. Huang S: **Non-genetic heterogeneity of cells in development: more than just noise**. *Development* 2009, **136**:3853–3862.

58. Kauffman SA: **The origin of order**. Oxford University Press, Oxford, 1993.

59. Huang S: **The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology?** *Bioessays* 2012, **34**:149–157.

Systematic description of the importance of landscapes in the context of evolution.

60. Mojtahedi M, Skupin A, Zhou JX, Castano IG, Leong-Quong R, Chang H, Giuliani A, Huang S: **Binary cell fate-decision as high-dimensional critical state transition**. [date unknown], [no volume].

61. Wang J, Xu L, Wang E, Huang S: **The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation**. *Biophys. J.* 2010, **99**:29–39.

62. Zhou JX, Brusch L, Huang S: **Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model**. *PLoS One* 2011, **6**:e14752.

63. Wang J, Zhang K, Xu L, Wang E: **Quantifying the Waddington landscape and biological paths for development and differentiation**. *Proc. Natl. Acad. Sci. U. S. A.* 2011, **108**:8257–8262.

64. Sisan DR, Halter M, Hubbard JB, Plant AL: **Predicting rates of cell state change caused by stochastic fluctuations using a data-driven landscape model**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:19262–19267.

65. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, Yuan G-C: **Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape**. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**:E5643–50.

66. Kafri R, Levy J, Ginzberg MB, Oh S, Lahav G, Kirschner MW: **Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle**. *Nature* 2013, **494**:480–483.

Mathematical framework to deduce cell growth dynamics from flow cytometry snapshot data of single cells.

67. Tyson DR, Garbett SP, Frick PL, Quaranta V: **Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data**. *Nat. Methods* 2012, **9**:923–928.

68. Qian H, Hong Q: **Cooperativity in Cellular Biochemical Processes: Noise-Enhanced Sensitivity, Fluctuating Enzyme, Bistability with Nonlinear Feedback, and Other Mechanisms for Sigmoidal Responses**. *Annu. Rev. Biophys.* 2012, **41**:179–204.