CellPress

Cell Stem Cell

Perspective

Computational Stem Cell Biology: Open Questions and Guiding Principles

Patrick Cahan,[1,13,14,*] Davide Cacchiarelli,[2,13] Sara-Jane Dunn,[3,4,13] Martin Hemberg,[5,13] Susana M. Chuva de Sousa Lopes,[6,13] Samantha A. Morris,[7,13] Owen J.L. Rackham,[8,13] Antonio del Sol,[9,10,11,13] and Christine A. Wells[12,13]
[1]Institute for Cell Engineering, Department of Biomedical Engineering, Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
[2]Telethon Institute of Genetics and Medicine (TIGEM), Armenise/Harvard Laboratory of Integrative Genomics, Pozzuoli, Italy d Department of Translational Medicine, University of Naples "Federico II," Naples, Italy
[3]DeepMind, 14-18 Handyside Street, London N1C 4DN, UK
[4]Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Jeffrey Cheah Biomedical Centre, Puddicombe Way, Cambridge Biomedical Campus, Cambridge CB2 0AW, UK
[5]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK
[6]Department Anatomy and Embryology, Leiden University Medical Center, Einthovenweg 20, Leiden, the Netherlands
[7]Department of Developmental Biology, Department of Genetics, Center of Regenerative Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA
[8]Centre for Computational Biology and The Program for Cardiovascular and Metabolic Disorders, Duke-NUS Medical School, Singapore, Singapore
[9]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, Belvaux 4366, Luxembourg
[10]CIC bioGUNE, Bizkaia Technology Park, 801 Building, 48160 Derio, Spain
[11]IKERBASQUE, Basque Foundation for Science, Bilbao 48013, Spain
[12]Centre for Stem Cell Systems, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, VIC 3010, Australia
[13]These authors contributed equally
[14]Lead Contact
*Correspondence: patrick.cahan@jhmi.edu
https://doi.org/10.1016/j.stem.2020.12.012

SUMMARY

Computational biology is enabling an explosive growth in our understanding of stem cells and our ability to use them for disease modeling, regenerative medicine, and drug discovery. We discuss four topics that exemplify applications of computation to stem cell biology: cell typing, lineage tracing, trajectory inference, and regulatory networks. We use these examples to articulate principles that have guided computational biology broadly and call for renewed attention to these principles as computation becomes increasingly important in stem cell biology. We also discuss important challenges for this field with the hope that it will inspire more to join this exciting area.

THE INTERSECTION OF COMPUTATIONAL BIOLOGY AND STEM CELL BIOLOGY

Computational tools have aided our understanding of stem cells and development at least since 1952, when Alan Turing used computer simulations to explore reaction diffusion as an explanation of embryonic patterning (Turing, 1952). In the intervening decades, computational tools have become inextricably linked to stem and developmental biology with the advent of high-throughput technologies such as nucleic acid sequencing, and with the increased prevalence of modeling and simulation. In this perspective, we discuss four current topics that epitomize the intersection of stem cell biology and computational biology: cell typing, lineage tracing, trajectory inference, and regulatory networks. We use these examples to articulate general principles of how computational biology can be leveraged to meet the future needs of stem cell biology. In the final section we highlight other topics at the interface between stem cell biology and computational biology (an intersection that we refer to as computational stem cell biology [CSCB]) but that we do not discuss in depth due to space constraints.

CURRENT CHALLENGES AND FUTURE OPPORTUNITIES

Cell Typing and Assessing Fidelity of Cell Fate Engineering

Three specific aims in CSCB relate to the concept of cell identity and are linked by the fact that they can be achieved with related computational methods that operate on molecular profiles (Figure 1). These are concerned with establishing (1) whether a cell population is pluripotent, (2) whether an individual cell is multipotent or pluripotent, and (3) the fidelity of engineered cell types when compared to their *in vivo* counterparts.

Prospective experimentation, such as a blastocyst complementation (Bradley et al., 1984), may be the gold standard for ascertaining whether a cell population is pluripotent, but it is not always possible nor desirable for all species (Müller et al., 2010). There are many cases where molecular surrogates of pluripotency are

 © 2020 Elsevier Inc.

CellPress

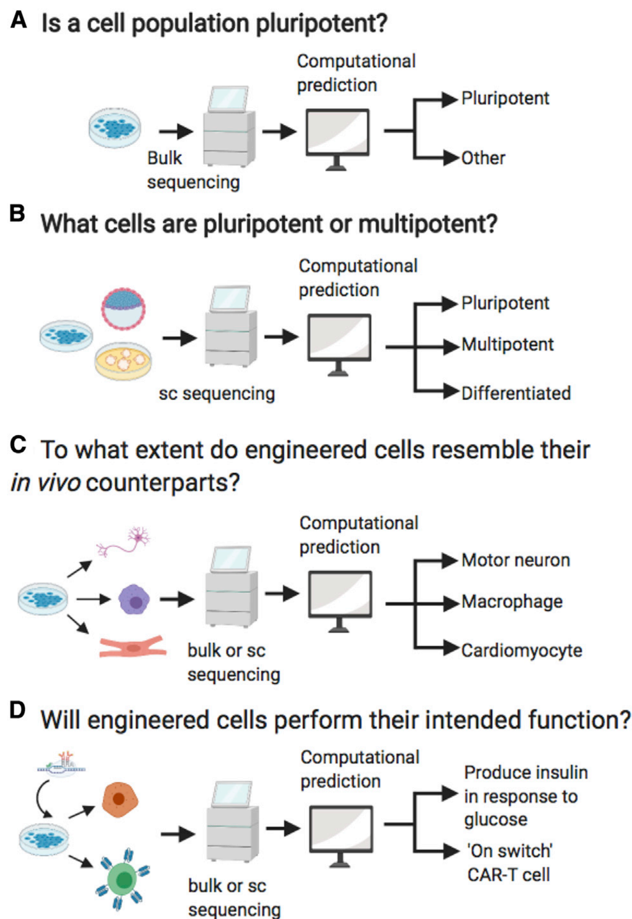

Figure 1. Computational Determination of Population and Single-Cell State, Identity, and Function from Single-Cell and Bulk Molecular Profiles

(A) First-generation CSCB methods to predict whether a population of cells was pluripotent or not.

(B) Second-generation methods extended these approaches from bulk population to single-cell level of resolution and allowed for prediction of multipotency.

(C) Methods to quantitatively measure the similarity of engineer cells (i.e., those derived by directed differentiation or direct conversion) to their *in vivo* counterparts.

(D) A pressing current challenge is to determine from a molecular profile the extent to which an engineered cell or cell population will behave as intended.

required, and traditionally these surrogates have been limited to a handful of markers, e.g., Pou5f1/Oct4 in pluripotent stem cells (Osorno et al., 2012) or Lgr5 in intestinal stem cells (Barker et al., 2007). However, relying on a small number of markers is unlikely to yield a robust outcome, for example in cases of partial reprogramming (Knaupp et al., 2017) or parthenote-derived hESCs (Sagi et al., 2019). Therefore, computational techniques that use genome-wide data to address this question have emerged. One of the first was Pluritest, which compared microarray-generated gene expression profiles of cell lines to a compendium of ESCs and differentiated cell types (Müller et al., 2011). The Pluritest approach leveraged non-negative matrix factorization (see glossary in Box 1) (Lee and Seung, 1999) to determine cell type conditional models of expression, which were then used as a basis to compute pluripotency scores on external samples. Subsequently, CellNet addressed the question of pluripotency by using a random forest classifier trained on gene regulatory networks extracted from a curated set of microarray expression data (Cahan et al., 2014; Morris et al., 2014). These and other "first generation" methods including ScoreCard, TeratoScore, and KeyGenes were trained on bulk data, which has well-known limitations relative to single-cell data (Avior et al., 2015; Bock et al., 2011; Roost et al., 2015). Profiling groups of cells may yield a low pluripotency score if some cells remained pluripotent but others had differentiated, and this would be indistinguishable from cultures where all cells had differentiated. Moreover, bulk data obscures the detection of distinct sub-states of pluripotency more easily discernible at the single-cell level (Neagu et al., 2020). Therefore, the second generation of computational pluripotency predicting methods relies on single-cell genome-wide data, mostly scRNA-seq. Currently, there are two approaches to do so.

One approach to identify stem cells from scRNA-seq data is to search for higher level features of stem cells in the query data. For example, stem cells tend to sporadically express lineage regulators, a phenomenon called priming in hematopoietic stem cells (Hu et al., 1997; Månsson et al., 2007; Miyamoto et al., 2002). Expression heterogeneity can be quantified by borrowing a core notion from information theory—that of entropy, which quantifies the degree of uncertainty associated with a random variable. Several methods have used this concept, including SLICE, which infers potency based on entropy of GO-defined gene sets, and SCENT, which infers potency based on the entropy of signaling networks, and this concept has been extended to quantify the maturation of pluripotent stem cell (PSC)-derived cardiomyocytes (Guo et al., 2017; Kannan et al., 2020; Teschendorff and Enver, 2017). A central property of stem cells is that they differentiate into distinct lineages. Therefore, if lineage relationships can be gleaned from single-cell data, then this information can be factored into a metric of fate potential. In a subsequent section, we describe trajectory inference (TI) methods that perform precisely this task. StemID and FateID are examples of potency prediction methods that count the number of lineages predicted to emerge from each cluster of single cells (Grün et al., 2016; Herman et al., 2018). They then incorporate this information with transcriptional entropy to produce a potency score. These methods used *a priori* determined features of stem cells (e.g., transcriptional heterogeneity, number of predicted descent lineages), whereas a more recent approach, CytoTRACE, took a hypothesis-free approach to find features that tracked with fate potency across a range of well-annotated scRNA-seq datasets (Gulati et al., 2020). Somewhat surprisingly, the number of genes expressed in a cell was discovered as the top predictor. Based on the rapid pace of innovation in this area, we expect that there will continue to be innovative approaches to derive and combine genome-wide features to predict cell fate potency.

The second class of approaches to predict cell fate potency is analogous to the first generation based on bulk data in that they compare gene expression from a query dataset to a reference dataset in which the stem and differentiated cells have been annotated. However, these comparisons are challenging due to differences in experimental conditions that may introduce batch effects that obscure meaningful comparisons. Although more

Box 1. Glossary of Terms

Auto-regulatory loop: network motif in which nodes positively regulate themselves and each other
Boolean network: a network in which each node can take on only binary values (ON/OFF), which is determined by the set of edges into that node and a defined regulatory function (e.g., OR)
Cell fate engineering: deriving a cell population of desired identity through directed differentiation of stem cells or through direct conversion
Entropy: in information theory, entropy is the uncertainty associated with the outcome of random variable
Maximum likelihood: a technique of estimating the parameters of a model such that the probability of the observed data is at the maximum
Maximum parsimony: a technique to reconstruct a phylogenetic tree by minimizing the number of nucleotide changes required to build the tree
Mutual information: in information theory, MI can be considered the amount of information about one random variable is gained by determining the value of another random variable
Network motif: sub-graphs that occur frequently in a network (e.g., positive feedback)
Non-negative matrix factorization: a group of matrix factorization approximation algorithms
Random forest: a machine learning classification technique that is based on sets of decision trees generated by repeated sampling from the training data
Regulon: the set of genes that are transcriptionally regulated by a transcription factor
Simpson's paradox: the appearance of trends in grouped data that disappear or change when examining the groups individually

standardized protocols for sample preparations may ameliorate these issues, the need for specialized computational approaches that go beyond standard clustering methods will remain (Ding et al., 2020; Ziegenhain et al., 2017). Already there are several methods, for example scmap, scID, singleCellNet, scPred, scClassify, Cell-BLAST, Moana, Garnett, and CellAssign (Alquicira-Hernandez et al., 2019; Boufea et al., 2020; Cao et al., 2020; Kiselev et al., 2018; Lin et al., 2019; Pliner et al., 2019; Tan and Cahan, 2019; Wagner and Yanai, 2018; Zhang et al., 2019) that perform "cell typing," recently reviewed by Abdelaal et al. (2019). These methods are also emerging as a way to address the third aim in this section, which is to determine the extent to which the products of cell fate engineering (CFE) resemble their "natural" counterparts.

Most methods dealing with predicting cell type and cell potency have adhered to best practices of scientific computational tools (e.g., availability of code, data, and documentation to ensure reproducibility and to enable improvements and extensions). Two areas that still warrant improvement are guidance on how to optimize parameters and how to assess the confidence of the outputs. For example, CellNet returns a classification score that represents the fraction of decision trees that predicted that the sample was a given cell type, but the associated sensitivity and precision of this score are not easily extractable. Pluritest and KeyGenes return the distribution of scores for training data so that the user can visualize the similarity of their samples to true ESCs and the predicted *in vivo* human organ/tissue counterparts, respectively. While ideally these outputs would be coupled to performance metrics such as sensitivity and precision, they are nonetheless sufficient to generate reliable hypotheses that have been experimentally tested in several systems (Takasato et al., 2015).

There are several issues in predicting cell type and cell potency that warrant special attention and focused effort. First, most current cell-typing methods have been developed and benchmarked primarily for terminally differentiated cell types whereas, by definition, stem and progenitor cells are uncommitted. Therefore, a first step will be to generate and use appropriate training data, such as scRNA-seq of embryonic and fetal sources. We note that because cells often exhibit a continuum of expression states as they differentiate (Sharma et al., 2020; Stumpf et al., 2017), supervised machine learning approaches to predict developmental staging and cell maturation will require very large training datasets.

Second, these methods need to incorporate other sources of molecular data to produce better predictors of cell identity and cell potency. Just as bulk profiling approaches have obscured the heterogeneity of cell phenotypes, current approaches to pathway or cell ontologies lack sufficient molecular resolution to derive cell-type-specific annotations. Therefore, single-cell integration of multiple molecular layers is needed to describe and understand cell state transitions. In addition to scATAC-seq, emerging technologies include the dual profiling of RNA and open chromatin, RNA and protein surface markers, and RNA and DNA methylation, as reviewed in Hu et al. (2018). The development and support of well-curated databases that link high dimensional genetic, epigenetic, transcriptomic, and proteomic data to specific phenotypes, such as tumorigenicity, immunogenicity, genome stability, editability, pluripotency, and differentiation potential (e.g., EBI-EMBL Expression atlas, scExpression atlas, Stemformatics, as well as hPSCreg) will enable new prediction algorithms (Choi et al., 2019; Papatheodorou et al., 2020). Principled methods that tie these molecular readouts together will be useful beyond more finely resolved predictions of cell identity and cell potency. For example, methods that can take into account the underlying genetic variability and the role of specific variants in determining cell fate can elucidate cellular mechanisms and guide future experiments (van der Wijst et al., 2020).

Third, we need to develop methods and a corresponding nomenclature to handle cases where CFE yields cell types that do not exist or have not been detected in a developmentally anchored reference atlas (Kime et al., 2019; Tonge et al., 2014;

CellPress

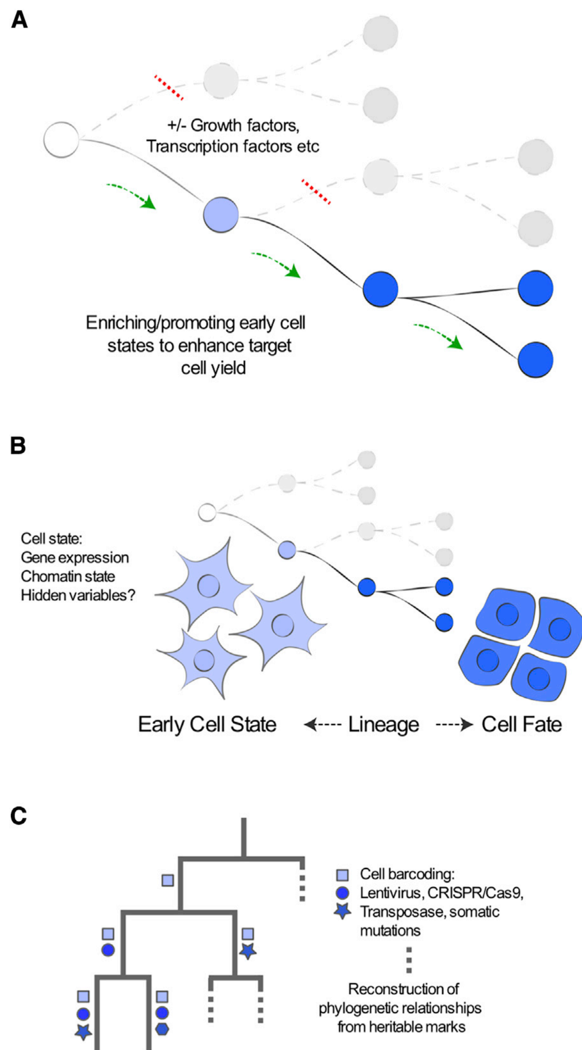

Figure 2. Single-Cell Lineage Tracing Applications
(A) Understanding how early state relates to eventual fate is valuable to design efficient cell engineering strategies. Enriching for or promoting desirable early cell states can enhance the yield of target cell types.
(B) Via the simultaneous capture of lineage and cell identity across a differentiation/reprogramming process, early cell state can be linked to eventual cell fate. These lineage tracing strategies indicate the existence of heritable properties that guide fate determination. However, eventual cell fate cannot be predicted based only on gene expression. Additional information, from chromatin accessibility assays such as single-cell ATAC-seq, and other "hidden variables" may serve a valuable role in uncovering these heritable properties.
(C) Schematic of cell labeling to enable phylogeny construction. Heritable cell labels are introduced via a variety of experimental methods, or naturally occurring somatic mutations can be exploited. Accumulation and inheritance of labels is used to reconstruct phylogenetic trees.

Yang et al., 2019a). Imperfectly differentiated or reprogrammed cells exist in hybrid and often transient states and the extent of artificial cell types has not been systematically explored (Morris et al., 2014), but it is a phenomenon that will negatively impact attempts to deconvolute population composition from bulk-derived data (Burke et al., 2020). Therefore, more expansive cell-typing methods that recognize hybrid identities and that find the specific attributes shared between the engineered cells and *in vivo* cell types are needed. Along these lines, we should develop methods to predict the behavior or function of cells that have been produced with synthetic biology—in other words, cells that are not intended to mimic an *in vivo* identity, but rather are engineered to perform specific functions (Figure 1; Del Vecchio et al., 2017; Qian et al., 2017; Toda et al., 2018, 2019).

Finally, we will need to capitalize on *in situ* technologies, which can include information such as cell morphology, structural characteristics, and the behavior of neighboring cells. These data can be used to perform more refined cell typing and to distill information about the specific contributors of niche elements to stem cell maintenance and behavior. Ultimately, we predict that *in situ* data will be used to characterize the fidelity of engineered tissues and organs on the combined basis of cell identity, cell type proportions, and localization.

Lineage Tracing

Whereas cell typing is concerned with cell identity, the central goal of lineage tracing is to identify all progeny that arise from an individual cell. Understanding lineage history is powerful in directing cell fate decisions to the desired outcome in the context of stem cell biology. For example, analyzing lineage tracing of alveolar epithelial type 2 (iAEC2) differentiation with a Continuous State Hidden Markov Model (Lin and Bar-Joseph, 2019) led to predictions that ultimately improved directed differentiation of iAEC2 cells (Figure 2A; Hurley et al., 2020). Although many lineage tracing approaches can achieve high spatial resolution, they typically do not capture many molecular features of each cell under observation, and thus provide limited insight into the regulators and mechanisms of cell fate decisions. To overcome this limitation, single-cell transcriptomics has recently taken a core role, fueling a new wave of lineage tracing tools. These methods are based on uniquely labeling individual cells by leveraging naturally occurring somatic mutations (Leung et al., 2017; Lodato et al., 2015) or experimentally introducing heritable genetic makers (Kester and van Oudenaarden, 2018). While genetic labeling or barcoding offers theoretically unlimited diversity to track cells and their progeny within a defined population (Lu et al., 2011; McKenna et al., 2016; Porter et al., 2014; Sun et al., 2014), these methods were originally incompatible with scRNA-seq because they relied on DNA sequencing. In response, a suite of tools enabling barcode expression as transcripts has allowed the capture of lineage information in parallel with single-cell transcriptomes (Alemany et al., 2018; Biddy et al., 2018; Bowling et al., 2020; Frieda et al., 2017; Raj et al., 2018; Spanjaard et al., 2018; Tusi et al., 2018; Wagner et al., 2018; Yao et al., 2017).

Viral barcoding for clonal analysis and lineage tracing has been deployed across several stem cell differentiation and reprogramming paradigms. The accessibility and malleability of cells within these systems renders them amenable to multiple rounds of viral transduction, representing a more tractable approach for cell barcoding, relative to genome editing. Furthermore, cells can be sampled throughout differentiation or reprogramming, enabling progenitor states to be linked to eventual fate. This strategy, coupled with two computational methods (logistic regression and a multilayer perceptron neural network), was used to determine how well gene expression state in progenitors accounts for eventual cell fate in hematopoiesis (Weinreb et al.,

2020). Later-stage sister cells were used to predict the dominant differentiation outcome, where gene expression at the later stage held greater predictive power than the expression state of the progenitors. These observations indicate the existence of heritable properties that guide fate determination that were not detected by scRNA-seq alone. Similar observations were made in transcription factor-mediated direct reprogramming of mouse embryonic fibroblasts (MEFs) to induced endoderm progenitors (iEPs) (Biddy et al., 2018). In contrast to these observations, an *in vivo* state-fate system uncovered distinct transcriptional states in HSCs that predicted differentiation capacity (Pei et al., 2020). One possible explanation for this apparent discrepancy is that the molecular states of *ex vivo* and *in vitro* progenitors are less well defined than their *in vivo* counterparts on a transcriptional versus epigenomic basis. Additional information from chromatin accessibility assays such as ATAC-seq and single-cell ATAC-seq will be invaluable to uncovering these heritable properties and in resolving these apparent discrepancies (Figure 2B). For example, an application of machine-learning and gene regulatory network analysis of gene expression and chromatin accessibility information showed that lineages that failed to convert to iEPs did so because the reprogramming transcription factors (TFs) were unable to properly regulate genes essential to iEP function and identity (Kamimoto et al., 2020). These observations were consistent with prior work suggesting that reprogramming to pluripotency is inefficient due to the required target genes being "locked" in heterochromatin and thus unavailable for targeting by reprogramming factors (Soufi et al., 2012). Collectively, this combination of lineage tracing, scRNA-seq and computational sleuthing has implicated factors determining chromatin accessibility as the likely heritable properties that influence cell differentiation and reprogramming outcome.

The computational analysis of single-cell lineage tracing is in its infancy and is currently facing several unmet needs. First, particularly for CRISPR-Cas9 single-cell lineage tracing, broadly applicable methods for phylogenetic tree reconstruction based on maximum-likelihood and maximum parsimony approaches are emerging (Figure 2C; Feng et al., 2019; Jones et al., 2020) but will need continuous development to accommodate the diversifying experimental lineage tracing toolbox. Indeed, novel computational methods are overcoming challenges presented by these complicated experimental platforms. For example, one computational method exploits the stochasticity of cell fate choice in development to overcome the requirement for multiple rounds of labeling to infer phylogenies (Weinreb and Klein, 2020). Second, as above, the computational integration of diverse data modalities will be critical to uncover the heritable properties that determine cell fate. Third, new techniques to visualize lineage in concert with state manifolds will be essential to fully interpret ground truth lineage data and reveal the existence of hidden variables that could be leveraged to fully understand the underlying molecular mechanisms that control the specification and maintenance of cell identity.

Trajectory Inference

The molecular state of a cell in a tissue or population is rarely static, but rather varies stochastically in response to its environment, and to reflect the stage of biological processes that animate it, such as cell cycle and differentiation. Sampling a population of cells by scRNA-seq or other single-cell molecular profiling captures this heterogeneity. Trajectory inference (TI) is the computational task of determining the position of single cells on temporally regulated biological processes. TI is powerful in many ways. First, it enables the identification of new transition stages or branch points, as well as stage-specific markers that can be used to prospectively isolate transient populations. Second, TI allows for the identification of clusters of genes correlated in temporal expression, and thereby allows for the determination of the function of unannotated genes via "guilt-by-association." Third, by placing cells in pseudotemporal order, TI allows for the inference of causal regulatory relationships and thus can identify regulators of differentiation and the subsequent cascade of transcriptional events. In practice, TI has most often been used to study differentiation and therefore is akin to lineage tracing in that it explores the lineage relationships. However, trajectories and lineages are fundamentally distinct measurements, with the latter requiring the experimental mapping of cell lineage relationships. Therefore, differences in the methods (inference based on transcriptional similarity versus molecular fingerprinting) and the timescales (hours to days versus days to weeks or longer) have entailed distinct computational approaches. In this section, we give a taste of how TI methods work, and we describe some early, pioneering applications of TI in the context of differentiation and CFE. Then, we touch on more recent advances in this area, discuss existing areas that require improvement, and highlight opportunities for conceptual and methodological advancement that will help TI methods to reach their potential.

Among the first TI methods were Wanderlust and Monocle, which were designed for mass cytometry and scRNA-seq data, respectively (Bendall et al., 2014; Trapnell et al., 2014). Key aspects of both of these approaches have endured as common features of most subsequent TI methods: (1) embed single-cell data in a lower dimensional space to provide a more efficient representation of the cells and a basis for a more biologically relevant cell-cell distance metric; (2) create a graph that links cells or group of cells; (3) infer the trajectory based on the topology of the cell-cell graph; and (4) place cells on the inferred trajectories, which is also referred to as determining pseudotime. The application of Monocle to differentiating myoblasts revealed the power of the TI approach in several ways. First, it identified transcriptional regulators of differentiation based on enrichment of TF binding sites in genes sharing the same temporal expression pattern. Second, it uncovered alternative, unexpected differentiation trajectories. Finally, it led to the identification and experimental validation of novel regulators of branch point decisions. Another early and pioneering application of TI was to determine the extent to which directed differentiation and direct programming to motor neurons follow the same developmental path (Briggs et al., 2017). Although no formal TI method was used in this study, trajectories and the relative ordering of cell states were inferred from dimensionality-reduced scRNA-seq data. Comparing directed differentiation and direct programming led to the observation that direct programming skipped intermediate stages characterized by the expression of patterning genes, the activation of which was subsequently recovered close to the final motor neuron stage. This study

CellPress

was a potent demonstration of how TI analysis can be used to explore fundamental questions, in this case of convergent development and the concept of cell types as attractors states.

Dozens of TI methods have since been invented that vary primarily in the analysis steps outlined above and in their assumption of the topology of the trajectory of the data (Saelens et al., 2019). For example, diffusion pseudotime uses a nonlinear dimension reduction technique that better reflects the continuous and noisy nature of differentiation than do linear methods such as PCA and ICA (Haghverdi et al., 2016). Based on benchmarking analyses, it is clear that there is no single best practice or best method, and therefore we expect that further refinements to this core pipeline will continue. More recently, however, very creative approaches that go beyond the core TI pipeline to predict differentiation paths from scRNA-seq have emerged. For example, RNA Velocity and subsequent extensions use the ratio of spliced to unspliced mRNA to model transcriptional kinetics and thereby predict the future state of a cell (Bergen et al., 2020; La Manno et al., 2018). One of the benefits of RNA Velocity is that it gives direct and automatic indication of the starting and terminal points of a trajectory. Another new approach is Waddington optimal-transport (WOT), which, unlike most TI methods, is robust to the underlying topology (e.g., linear, branching, cycle) of the developmental trajectory (Schiebinger et al., 2019). WOT employs the mathematical tool of optimal transport to identify cell descendants and ancestors among coupled time points using proliferation- and apoptosis-related gene expression. WOT assumes that cell proliferation is the driving force of developmental processes connecting cell descendants to ancestors along a single trajectory. Another innovation of WOT was the use of ligand-receptor interaction analysis to uncover crosstalk along trajectories and to identify crucial interactions that orchestrate cell fate decisions at branch points during reprogramming.

There are several major open questions and opportunities for advancement in TI. First, because TI methods are so new, they generally do not have strong guidance on how to determine whether optimization is required, and if so, how to perform this optimization. Second, many TI methods rely on user input to identify starting and terminal points, but this is often not known. The integration of other methods such as CytoTRACE or RNA Velocity to determine these starting and end points in a systematic way would be valuable. Third, most methods do not handle disjoint sets of cells well (Cao et al., 2019). Fourth, the assumption that transcriptional resemblance equates to lineage relation might not hold across all developmental contexts as different sources of expression variation could mask the bona fide developmental trajectory. Therefore, approaches to identify such cases should be devised. Fifth, although gold standards to evaluate and compare TI methods have matured, the *in vivo* data still rely largely on temporal ordering as defined by original publications rather than orthogonal data. Fate-state data are an important dimension to add to growing TI gold standards. Sixth, as cell fate decisions are governed by spatial-temporal signaling, computational methods that infer paracrine signaling during development from *in situ* sequencing data are needed. Finally, especially in the context of comparing *in vitro* with *in vivo* development, it would be valuable to have formal methods to systematically compare trajectories in a quantitative manner.

Inferring and Using Regulatory Networks

Critical to understanding cell identity and its emergence via developmental trajectories is to understand the regulatory networks that drive cellular decision making. Such networks involve many different biological macromolecules (e.g., genes, proteins, metabolites, and non-coding RNA) that interact in diverse ways to produce the incredible diversity of emergent behavior we see across each of our cell types. For instance, specific configurations of these networks control differentiation (Simões-Costa and Bronner, 2015) or direct responses to external stimuli (Bourret and Stock, 2002). Computational representations of regulatory networks enable the modeling of cell behavior as a system and the prediction of how cell state changes over time or upon perturbation. In the context of stem cell biology, regulatory networks have been used to identify key mediators of pluripotency, multipotency, self-renewal, and cell fate decisions, to name just a few (Chen et al., 2008; Nishiyama et al., 2009; Yachie-Kinoshita et al., 2018; Zhou et al., 2007). Moreover, networks provide a formalism for reasoning about general mechanisms that explain stem cell properties. For example, network concepts have been used to explain cell behavior like cell state stability, cell fate choices, and the effect of targeted perturbations (Dunn et al., 2014; Zhou et al., 2011). In this section, we describe the most prevalent experimental and computational methods used to identify and analyze regulatory networks. We briefly mention several seminal applications in stem cell biology, development, and CFE. We finish the section by discussing how single-cell approaches are revolutionizing this area and the pressing questions that need to be addressed in the future.

There are several classes of networks used to explore and understand cellular decision making. Arguably, the most frequently used in stem cell biology is the transcriptional network, commonly referred to as a gene regulatory network (GRN) (Erwin and Davidson, 2009; Karlebach and Shamir, 2008). In GRNs, nodes represent genes while edges represent transcriptional regulation and are present only between genes encoding TFs and their predicted target genes, which may include other TFs. Edges can be inferred one TF at a time using ChIP-ChIP or ChIP-seq (Figure 3A). In fact, a pair of pioneering applications of ChIP-ChIP identified the regulons of the core pluripotency regulators Oct4, Sox2, and Nanog (Boyer et al., 2005; Loh et al., 2006). In both mouse and human ESCs, these factors form an auto-regulatory loop, a type of network motif, which helps to explain how transient loss of expression of any one factor can be tolerated (Chen et al., 2008). Later work dissecting other targets of these core factors revealed that another network motif, mutual inhibition of targets of these factors, contributes to early lineage commitment (Loh and Lim, 2011; Thomson et al., 2011). A benefit of defining edges with ChIP-seq is that it yields transcription factor binding sites (TFBSs), which can subsequently be used to search for TF activity in other experimental settings (e.g., enrichment in gene clusters or in accessible genomic regions identified via DNase-hypersensitivity or ATACseq). Edges can also be inferred by transcriptional profiling after modulation of TF expression. For example, hundreds of mouse and human PSC lines have been engineered such that induction of a single TF is drug controllable. These lines have been used to identify the transcriptional consequences of induction of each of hundreds of TFs (Nakatake et al., 2020; Nishiyama et al., 2009).

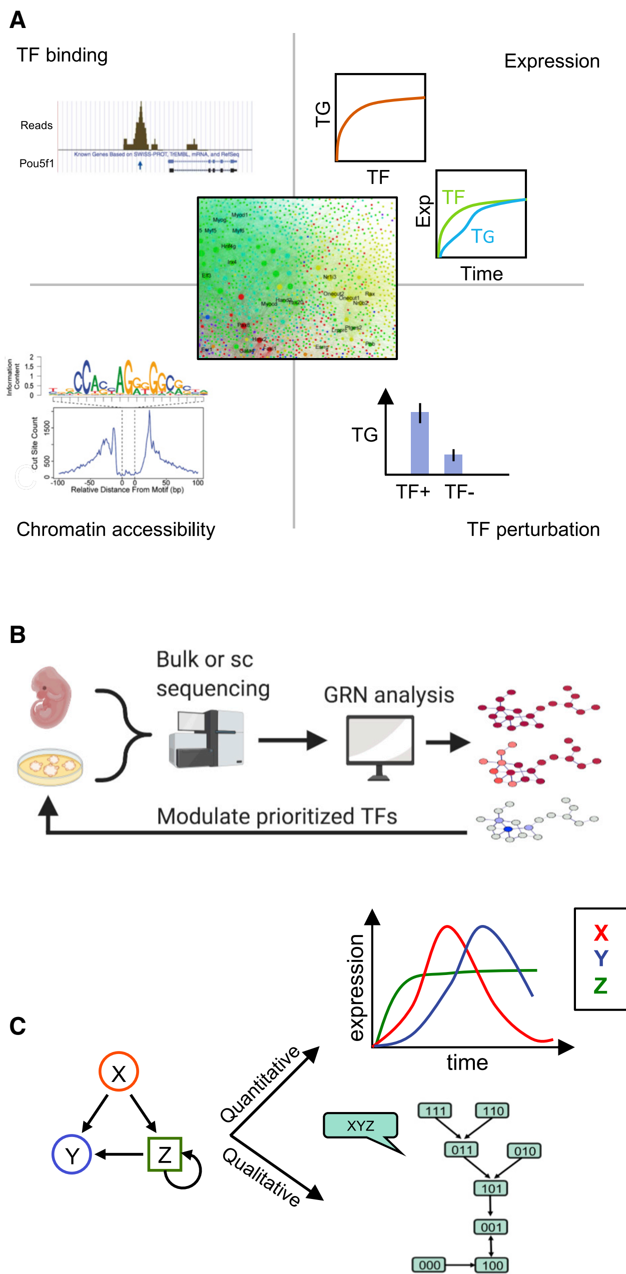

Figure 3. Derivation and Application of Cellular Networks
(A) Experimental strategies to reconstruct transcriptional regulatory networks. Top-left: ChIP-ChIP and ChIP-seq identify transcription factor binding sites. Top right: Association in expression patterns between TF and putative target genes across perturbations, or in a time-lagged manner, imply a regulatory relationship. Bottom left: TF motif co-localization with TG regulatory sequence. Bottom right: TF modulation impacts TG expression.
(B) Leveraging networks to improve CFE. GRNs are first constructed from *in vivo* data. Computational integration of the activity of network components in engineered cells is used to predict regulators to modulate to improve the CFE outcome.
(C) Generative networks can be used for either quantitative (using ordinary differentiation equations) or qualitative (using Boolean networks) dynamic simulations.

Recently, this concept has been extended to the single-cell level with approaches such as Perturb-Seq or Reprogram-Seq, which can in parallel test the effect of many TFs in isolation or combination (Dixit et al., 2016; Duan et al., 2019). Groups have compiled the results of these and other experimental approaches into large databases of predicted regulatory relationships (or edges) (Gheorghe et al., 2019; Szklarczyk et al., 2019), which can then be used to perform network analysis in other contexts or as benchmarks against which to evaluate the performance of GRN inference methods.

Another class of methods to identify GRNs is based on the premise that a variety of perturbations will elicit TF-target associations in expression. By computing the pairwise association in expression between TFs and all genes across a range of perturbations or states, it is possible to detect regulatory relationships (Le Novère, 2015). There are many such network reconstruction methods that differ in the metric of association (e.g., Pearson correlation or Mutual information [Wang and Huang, 2014]), approaches to deal with the preponderance of false positives that come from indirect interactions (Margolin et al., 2006), whether the method predicts direction of regulation (promotes or represses), and whether the method leverages TFBSs or epigenomic information, among many other aspects. These methods have been subject to community-driven benchmarking, which has led to the conclusion that when applied to eukaryotic systems, any single co-expression based method has a qualitatively low performance, but aggregating results across methods yields better predictions (Marbach et al., 2012). An oft-maligned culprit is the difficulty in distinguishing direct from indirect effects, leading to a high number of false positives. It was initially anticipated that single-cell data would improve GRN reconstruction by avoiding Simpson's paradox and thus reducing the false positive rate (Trapnell, 2015). However, initial benchmarking of single-cell GRN reconstruction methods suggest that single-cell data alone does not improve GRN reconstruction performance (Chen and Mar, 2018; Pratapa et al., 2019). Nonetheless, reconstructing GRNs from single-cell data has yielded important insights, including pinpointing the roles of Sox and Hox TFs in the emergence of hematopoietic lineages from mesoderm (Moignard et al., 2015), identifying the targets of pluripotency TFs as mESCs transit from naive to primed to neurectoderm (Chan et al., 2017; Stumpf et al., 2017) and identifying novel and specific regulators of outflow tract cell differentiation (de Soysa et al., 2019). There are many directions for exploration and improvement of GRN methods from single-cell data; one logical next step is to determine how to optimally incorporate pseudo-temporal information both to infer causal relations and to derive dynamic networks (Qiu et al., 2020).

One of the most potent applications of regulatory networks has been in the area of CFE, to predict how to control cell state and fate transitions. The first such algorithms, such as CellNet, Mogrify, and SeeSawPred, relied on bulk genomic profiling data and used cell-type-specific GRNs to identify candidate TFs whose expression could be modulated to engineer desired cell fate transitions (Figure 3B; Cahan et al., 2014; Hartmann et al., 2018; Rackham et al., 2016). These TFs may not have been identified by a more general differential expression analysis because they often exhibit modest differences in expression. However, GRN-based approaches that account for changes in regulon activity will detect fate-influencing TFs and thereby reduce the amount of expensive and time-consuming experimental work required. The advent of single-cell approaches has opened the possibility of inventing improved CFE

computational methods or refining them to target the ever-growing number of cell types and states. One such scRNA-seq-based GRN-based method identified a combination of small molecules that increased reprogramming efficiency by inferring and analyzing the GRN governing the corresponding reprogramming trajectory (Tran et al., 2019). Another application of scRNA-seq and GRNs used concepts from information theory to identify synergistic factors that enabled the conversion of hindbrain neuroepithelial cells into medial floor plate midbrain progenitors (Okawa et al., 2018). These are among the first of what will undoubtedly be many approaches and examples of CFE computational methods that leverage single-cell data.

There are many other ways that GRNs can be used to explore and learn about stem cells. An example of a different application of GRNs is the Reasoning Engine for Interaction Networks (RE:IN), which offers an extension of the Boolean network formalism (Peter and Davidson, 2017) that allows for uncertainty in network topology (Dunn et al., 2014; Yordanov et al., 2016). RE:IN borrows a technique from the field of formal verification (Bartocci and Lió, 2016) to incorporate experimental observations as constraints on the trajectories that a valid network should produce. In this way, it enables the user to identify the set of GRNs consistent with their experimental observations, and subsequently to use these candidate GRNs to explore the dynamic behavior of their system (Figure 3C). Exemplifying this, RE:IN has been used to guide experimental validation of untested, predicted behavior, revealing how the naive state in mouse is sustained or lost via non-trivial interactions between key pluripotency factors, how a dynamic, evolving network of interacting phosphatases regulates commitment and differentiation in the interfollicular epidermis, and which genetic perturbations accelerate and enhance the efficiency of reprogramming (Dunn et al., 2019; Mishra et al., 2017).

Beyond adapting CFE computational methods for single-cell data, there are several crucial, unanswered questions in this area. First, how do we optimally use GRNs to identify the most efficient ways to elicit an intended cell fate conversion? Concepts from network theory and control theory will be helpful to address this question, as will simulation systems that direct *in silico* screens. It is possible that the optimal use of network analysis will depend on the conversion under study and the nature of the given network itself. Second, how accurate and complete must GRNs be to engineer cell fate with a specified precision? Answering this fundamental question will also require a combination of theory and simulation, and in doing so will allow us to evaluate the practical utility of GRN reconstruction methods. Third, how can dynamic GRNs be leveraged to account for temporal dependence of expression states? In other words, CFE computational methods should not just predict what regulators to modulate, but in what order. Finally, how can GRNs be formally linked to signaling networks? Answering this last question will enable a new class of CFE computational methods that predict sets of small molecules, cytokines, and growth factors to enable cell fate conversions.

SYNTHESIS AND OUTLOOK

There are many other topics that sit at the intersection of stem cell biology and computational biology that we have not discussed due to space constraints. We mention some of them here to indicate the pervasiveness of computation in stem cell biology and to highlight some current unmet needs. A large, open challenge is to identify both coding and non-coding genetic variants that change the propensity of differentiation into specific lineages, which has implications in understanding both congenital disorders (Zhang and Lupski, 2015) and *in vitro* differentiation bias (Di Giorgio et al., 2008; Hu et al., 2010; Osafune et al., 2008). Another challenge is developing methods to leverage single-cell epigenomics, single-cell multi-omics (Macaulay et al., 2017), proteomics (van Hoof et al., 2012; Palii et al., 2019), phosphoproteomics (Kimura et al., 2020), and Hi-C (Di Stefano et al., 2020; Dileep et al., 2019; Kim et al., 2020; Zhang et al., 2020) to gain a more complete understanding of stem cells and their differentiation pathways. We note that some of these technologies have not yet reached single-cell resolution. Additionally, another type of biological network beyond GRNs are intercellular signaling networks (Yang et al., 2019b), which combine prior knowledge of ligand-receptor complexes with statistical frameworks to predict tissue-specific cell-cell communication networks that contribute to development and tissue homeostasis (Camp et al., 2017; Efremova et al., 2020; Raredon et al., 2019; Skelly et al., 2018; Vento-Tormo et al., 2018). In the future, predictive cell-cell interaction models will be able to inform key cell-cell interactions responsible for maintaining tissue homeostasis and supporting tissue regeneration. Moreover, the comparison of tissue-specific cell-cell interaction networks with cell-cell interactomes of pathological or injured tissues will identify dysregulated interactions that can guide the development of strategies to restore tissue homeostasis. The integration of these models of tissue-specific cell-cell interactions with imaging-based technologies for spatial transcriptome reconstruction (Halpern et al., 2017; Karaiskos et al., 2017; Rodriques et al., 2019) will enable the characterization of the complete interactome in a spatially resolved manner, and therefore enable the generation of more accurate predictions.

A final area that we did not discuss above is cell-based modeling, which is a technique that uses *in silico* representations to explore how cells interact and change over time (Ghaffarizadeh et al., 2018; Mirams et al., 2013; Sharpe, 2017). The long history of cell-based modeling of stem cells and development has included the exploration of alternative formulations of regulatory networks that spatially pattern the limb (Uzkudun et al., 2015) and the prediction of embryonic toxicity (Kleinstreuer et al., 2013). We anticipate that more accurate and powerful cell-based modeling systems will be generated by incorporating knowledge gained from single-cell sequencing and that such models will be used to iteratively design multi-cellular behavior and function.

In our exploration of how computational biology is being applied to stem cell biology, a set of core values or guiding principles emerged. While this list is not comprehensive, and in some ways it is applicable to computational biology more generally, it does include the most discussed and, we feel, most crucial principles that can help to meet the challenges of today and tomorrow:

1. Computational methods should have guidance that clearly describes how to interpret the significance or confidence of the method's results, and this guidance must go beyond

standard software documentation (Lee, 2018). Often, the output of computational methods is an ordered list of testable hypotheses, for example a list of transcription factors regulating a process. Since testing all of the possible hypotheses experimentally is often prohibitive, knowing how to interpret the significance of the method's results is crucial so that the user can prioritize experimental efforts on the most promising candidates.

2. Computational methods should provide clear guidance on when and how they should be optimized. Most computational biology tools have multiple parameters that impact analysis outcomes. While devising the methods, creators typically select default values based on optimization on their test data. A method's performance—for example sensitivity, accuracy, or execution time—can degrade when a user's data differ substantially from the test data used to derive the default parameters. Therefore, method creators should describe how to determine whether additional optimization is required.
3. Computational methods should be implemented and used in a way to ensure reproducibility. Data need to be described and stored so that they can be re-analyzed later to facilitate data integration and to allow for computational inference of phenotype. The widespread recognition of the benefits of FAIR data management (Wilkinson et al., 2016) and the fact that most journals now require deposition of genome-scale data in public repositories makes this principle widely adhered to. However, the varying degrees to which journals require code to be freely available has continued to hamper reproducibility of computational analytics (Papin et al., 2020).
4. We value tight, mutually beneficial collaborations between computationalists and experimentalists in which ideas and knowledge flow both ways. Such deep relationships will result in more productive studies because mathematical, statistical, and computational considerations are incorporated into experimental design, and because biological input from the experimentalists ensures that the computational models are relevant and incisive (Knapp et al., 2015). These types of deep interactions are also valued because they promote entry to the field by people from nontraditional backgrounds such as physics, economics, or other data sciences, and because they lead to the capture and standardization of metadata specific to stem cell biology. And finally, bringing together distinct sets of expertise can fuel new ideas and break away from entrenched norms.
5. Ideally, the initial publication of a method at the intersection of computational biology and stem cell biology will include a prospective, experimental assessment. However, we recognize that this is not always practical. In place of this validation, the publication should include a comparison to already published results or a discussion of how the method could be experimentally assessed. Such a discussion can lead to the development of standardized benchmarks that make the comparison and further improvement of a class of methods fairer and more efficient.

Adhering to these principles will help us to meet our challenges and goals in several ways. First, being tightly integrated with our experimental collaborators will make us more responsive to the wider stem cell field, while at the same time ensuring that our contributions are valued appropriately. Second, by setting a standard of reliable and interpretable methods, we will broaden the use of our tools and establish solid credibility in our field's work. Third, ensuring that our methods are accessible and modifiable will allow for their efficient improvement and adaptation for newly emerging questions and data-generating technologies. Finally, explicit consideration of prospective validation will help to ensure that methods are honed to address specific hypotheses.

ACKNOWLEDGMENTS

P.C. is supported by the National Institutes of Health (R35GM124725). D.C. was supported by Fondazione Telethon Core Grant, Armenise-Harvard Foundation Career Development Award, European Research Council (grant agreement 759154, CellKarma), and the Rita-Levi Montalcini program from MIUR.

DECLARATION OF INTERESTS

O.J.L.R is a co-founder, scientific advisory board member, and shareholder of Mogrify Ltd, a cell therapy company. D.C. is founder, shareholder, and consultant of Next Generation Diagnostic srl.

REFERENCES

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M.J.T., and Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol. *20*, 194.

Alemany, A., Florescu, M., Baron, C.S., Peterson-Maduro, J., and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. Nature *556*, 108–112.

Alquicira-Hernandez, J., Sathe, A., Ji, H.P., Nguyen, Q., and Powell, J.E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biol. *20*, 264.

Avior, Y., Biancotti, J.C., and Benvenisty, N. (2015). TeratoScore: Assessing the Differentiation Potential of Human Pluripotent Stem Cells by Quantitative Expression Analysis of Teratomas. Stem Cell Reports *4*, 967–974.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., and Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003–1007.

Bartocci, E., and Lió, P. (2016). Computational modeling, formal analysis, and tools for systems biology. PLoS Comput. Biol. *12*, e1004591.

Bendall, S.C., Davis, K.L., Amir, A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell *157*, 714–725.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414.

Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature *564*, 219–224.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.

Boufea, K., Seth, S., and Batada, N.N. (2020). scID Uses Discriminant Analysis to Identify Transcriptionally Equivalent Cell Types across Single-Cell RNA-Seq Data with Batch Effect. iScience *23*, 100914.

Bourret, R.B., and Stock, A.M. (2002). Molecular information processing: lessons from bacterial chemotaxis. J. Biol. Chem. *277*, 9625–9628.

Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.-C., Fujiwara, Y., Li, B.E., et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell *181*, 1410–1422.e27.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255–256.

Briggs, J.A., Li, V.C., Lee, S., Woolf, C.J., Klein, A., and Kirschner, M.W. (2017). Mouse embryonic stem cells can differentiate via multiple paths to the same state. eLife *6*, e26945.

Burke, E.E., Chenoweth, J.G., Shin, J.H., Collado-Torres, L., Kim, S.-K., Micali, N., Wang, Y., Colantuoni, C., Straub, R.E., Hoeppner, D.J., et al. (2020). Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. Nat. Commun. *11*, 462.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

Camp, J.G., Sekine, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., Kanton, S., Kageyama, J., Damm, G., Seehofer, D., et al. (2017). Multilineage communication regulates human liver bud development from pluripotency. Nature *546*, 533–538.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502.

Cao, Z.-J., Wei, L., Lu, S., Yang, D.-C., and Gao, G. (2020). Searching large-scale scRNA-seq databases via unbiased cell embedding with Cell BLAST. Nat. Commun. *11*, 3458.

Chan, T.E., Stumpf, M.P.H., and Babtie, A.C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst. *5*, 251–267.e3.

Chen, S., and Mar, J.C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. BMC Bioinformatics *19*, 232.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Choi, J., Pacheco, C.M., Mosbergen, R., Korn, O., Chen, T., Nagpal, I., Englart, S., Angel, P.W., and Wells, C.A. (2019). Stemformatics: visualize and download curated stem cell data. Nucleic Acids Res. *47* (D1), D841–D846.

de Soysa, T.Y., Ranade, S.S., Okawa, S., Ravichandran, S., Huang, Y., Salunga, H.T., Schricker, A., Del Sol, A., Gifford, C.A., and Srivastava, D. (2019). Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. Nature *572*, 120–124.

Del Vecchio, D., Abdallah, H., Qian, Y., and Collins, J.J. (2017). A blueprint for a synthetic genetic feedback controller to reprogram cell fate. Cell Syst. *4*, 109–120.e11.

Di Giorgio, F.P., Boulting, G.L., Bobrowicz, S., and Eggan, K.C. (2008). Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. Cell Stem Cell *3*, 637–648.

Di Stefano, M., Stadhouders, R., Farabella, I., Castillo, D., Serra, F., Graf, T., and Marti-Renom, M.A. (2020). Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. Nat. Commun. *11*, 2564.

Dileep, V., Wilson, K.A., Marchal, C., Lyu, X., Zhao, P.A., Li, B., Poulet, A., Bartlett, D.A., Rivera-Mulia, J.C., Qin, Z.S., et al. (2019). Rapid Irreversible Transcriptional Reprogramming in Human Stem Cells Accompanied by Discordance between Replication Timing and Chromatin Compartment. Stem Cell Reports *13*, 193–206.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

Duan, J., Li, B., Bhakta, M., Xie, S., Zhou, P., Munshi, N.V., and Hon, G.C. (2019). Rational reprogramming of cellular states by combinatorial perturbation. Cell Rep. *27*, 3486–3499.e6.

Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.

Dunn, S.-J., Li, M.A., Carbognin, E., Smith, A., and Martello, G. (2019). A common molecular logic determines embryonic stem cell self-renewal and reprogramming. EMBO J. *38*, 38.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat. Protoc. *15*, 1484–1506.

Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nat. Rev. Genet. *10*, 141–148.

Feng, J., DeWitt, W.S., McKenna, A., Simon, N., Willis, A.D., and Matsen, F.A. (2019). Estimation of cell lineage trees by maximum-likelihood phylogenetics. bioRxiv. https://doi.org/10.1101/595215.

Frieda, K.L., Linton, J.M., Hormoz, S., Choi, J., Chow, K.K., Singer, Z.S., Budde, M.W., Elowitz, M.B., and Cai, L. (2017). Synthetic recording and in situ readout of lineage information in single cells. Nature *541*, 107–111.

Ghaffarizadeh, A., Heiland, R., Friedman, S.H., Mumenthaler, S.M., and Macklin, P. (2018). PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLoS Comput. Biol. *14*, e1005991.

Gheorghe, M., Sandve, G.K., Khan, A., Chèneby, J., Ballester, B., and Mathelier, A. (2019). A map of direct TF-DNA interactions in the human genome. Nucleic Acids Res. *47*, e21.

Grün, D., Muraro, M.J., Boisset, J.-C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., van Es, J., Jansen, E., Clevers, H., et al. (2016). De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. Cell Stem Cell *19*, 266–277.

Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405–411.

Guo, M., Bao, E.L., Wagner, M., Whitsett, J.A., and Xu, Y. (2017). SLICE: determining cell differentiation and lineage based on single cell entropy. Nucleic Acids Res. *45*, e54.

Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nat. Methods *13*, 845–848.

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 352–356.

Hartmann, A., Okawa, S., Zaffaroni, G., and Del Sol, A. (2018). SeesawPred: A Web Application for Predicting Cell-fate Determinants in Cell Differentiation. Sci. Rep. *8*, 13355.

Herman, J.S., Sagar, and Grün, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat. Methods *15*, 379–386.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. (1997). Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. *11*, 774–785.

Hu, B.-Y., Weick, J.P., Yu, J., Ma, L.-X., Zhang, X.-Q., Thomson, J.A., and Zhang, S.-C. (2010). Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proc. Natl. Acad. Sci. USA *107*, 4335–4340.

Hu, Y., An, Q., Sheu, K., Trejo, B., Fan, S., and Guo, Y. (2018). Single Cell Multi-Omics Technology: Methodology and Application. Front. Cell Dev. Biol. *6*, 28.

Hurley, K., Ding, J., Villacorta-Martin, C., Herriges, M.J., Jacob, A., Vedaie, M., Alysandratos, K.D., Sun, Y.L., Lin, C., Werder, R.B., et al. (2020). Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. Cell Stem Cell *26*, 593–608.e8.

Jones, M.G., Khodaverdian, A., Quinn, J.J., Chan, M.M., Hussmann, J.A., Wang, R., Xu, C., Weissman, J.S., and Yosef, N. (2020). Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. Genome Biol. *21*, 92.

Kamimoto, K., Hoffmann, C.M., and Morris, S.A. (2020). CellOracle: Dissecting cell identity via network inference and in silico gene perturbation. bioRxiv. https://doi.org/10.1101/2020.02.17.947416.

Kannan, S., Farid, M., Lin, B.L., Miyamoto, M., and Kwon, C. (2020). Transcriptomic entropy quantifies cardiomyocyte maturation at single cell level. bioRxiv. https://doi.org/10.1101/2020.04.02.022632.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. Science *358*, 194–199.

Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. *9*, 770–780.

Kester, L., and van Oudenaarden, A. (2018). Single-Cell Transcriptomics Meets Lineage Tracing. Cell Stem Cell *23*, 166–179.

Kim, H.-J., Yardımcı, G.G., Bonora, G., Ramani, V., Liu, J., Qiu, R., Lee, C., Hesson, J., Ware, C.B., Shendure, J., et al. (2020). Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. PLoS Comput. Biol. *16*, e1008173.

Kime, C., Kiyonari, H., Ohtsuka, S., Kohbayashi, E., Asahi, M., Yamanaka, S., Takahashi, M., and Tomoda, K. (2019). Induced 2C Expression and Implantation-Competent Blastocyst-like Cysts from Primed Pluripotent Stem Cells. Stem Cell Reports *13*, 485–498.

Kimura, A., Toyoda, T., Iwasaki, M., Hirama, R., and Osafune, K. (2020). Combined Omics Approaches Reveal the Roles of Non-canonical WNT7B Signaling and YY1 in the Proliferation of Human Pancreatic Progenitor Cells. Cell Chem. Biol. *27*, 1–12.

Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. Nat. Methods *15*, 359–362.

Kleinstreuer, N., Dix, D., Rountree, M., Baker, N., Sipes, N., Reif, D., Spencer, R., and Knudsen, T. (2013). A computational model predicting disruption of blood vessel development. PLoS Comput. Biol. *9*, e1002996.

Knapp, B., Bardenet, R., Bernabeu, M.O., Bordas, R., Bruna, M., Calderhead, B., Cooper, J., Fletcher, A.G., Groen, D., Kuijper, B., et al. (2015). Ten simple rules for a successful cross-disciplinary collaboration. PLoS Comput. Biol. *11*, e1004214.

Knaupp, A.S., Buckberry, S., Pflueger, J., Lim, S.M., Ford, E., Larcombe, M.R., Rossello, F.J., de Mendoza, A., Alaei, S., Firas, J., et al. (2017). Transient and permanent reconfiguration of chromatin and transcription factor occupancy drive reprogramming. Cell Stem Cell *21*, 834–845.e6.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. Nat. Rev. Genet. *16*, 146–158.

Lee, B.D. (2018). Ten simple rules for documenting scientific software. PLoS Comput. Biol. *14*, e1006561.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

Leung, M.L., Davis, A., Gao, R., Casasent, A., Wang, Y., Sei, E., Vilar, E., Maru, D., Kopetz, S., and Navin, N.E. (2017). Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. Genome Res. *27*, 1287–1299.

Lin, C., and Bar-Joseph, Z. (2019). Continuous-state HMMs for modeling time-series single-cell RNA-Seq data. Bioinformatics *35*, 4707–4715.

Lin, Y., Cao, Y., Kim, H.J., Salim, A., Speed, T.P., Lin, D., Yang, P., and Yang, J.Y.H. (2019). scClassify: hierarchical classification of cells. bioRxiv. https://doi.org/10.1101/776948.

Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S., Chittenden, T.W., D'Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. Science *350*, 94–98.

Loh, K.M., and Lim, B. (2011). A precarious balance: pluripotency factors as lineage specifiers. Cell Stem Cell *8*, 363–369.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat. Biotechnol. *29*, 928–933.

Macaulay, I.C., Ponting, C.P., and Voet, T. (2017). Single-Cell Multiomics: Multiple Measurements from Single Cells. Trends Genet. *33*, 155–168.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., et al. (2007). Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity *26*, 407–419.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., and Stolovitzky, G.; DREAM5 Consortium (2012). Wisdom of crowds for robust gene network inference. Nat. Methods *9*, 796–804.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics *7* (*Suppl 1*), S7.

McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science *353*, aaf7907.

Mirams, G.R., Arthurs, C.J., Bernabeu, M.O., Bordas, R., Cooper, J., Corrias, A., Davit, Y., Dunn, S.-J., Fletcher, A.G., Harvey, D.G., et al. (2013). Chaste: an open source C++ library for computational physiology and biology. PLoS Comput. Biol. *9*, e1002970.

Mishra, A., Oulès, B., Pisco, A.O., Ly, T., Liakath-Ali, K., Walko, G., Viswanathan, P., Tihy, M., Nijjher, J., Dunn, S.-J., et al. (2017). A protein phosphatase network controls the temporal and spatial dynamics of differentiation commitment in human epidermis. eLife *6*, 6.

Miyamoto, T., Iwasaki, H., Reizis, B., Ye, M., Graf, T., Weissman, I.L., and Akashi, K. (2002). Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. Dev. Cell *3*, 137–147.

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

Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell *158*, 889–902.

Müller, F.-J., Goldmann, J., Löser, P., and Loring, J.F. (2010). A call to standardize teratoma assays used to define human pluripotent cell lines. Cell Stem Cell *6*, 412–414.

Müller, F.-J., Schuldt, B.M., Williams, R., Mason, D., Altun, G., Papapetrou, E.P., Danner, S., Goldmann, J.E., Herbst, A., Schmidt, N.O., et al. (2011). A bioinformatic assay for pluripotency in human cells. Nat. Methods *8*, 315–317.

Nakatake, Y., Ko, S.B.H., Sharov, A.A., Wakabayashi, S., Murakami, M., Sakota, M., Chikazawa, N., Ookura, C., Sato, S., Ito, N., et al. (2020). Generation and profiling of 2,135 human ESC lines for the systematic analyses of cell states perturbed by inducing single transcription factors. Cell Rep. *31*, 107655.

Neagu, A., van Genderen, E., Escudero, I., Verwegen, L., Kurek, D., Lehmann, J., Stel, J., Dirks, R.A.M., van Mierlo, G., Maas, A., et al. (2020). In vitro capture and characterization of embryonic rosette-stage pluripotency between naive and primed states. Nat. Cell Biol. *22*, 534–545.

Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y., et al. (2009). Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell *5*, 420–433.

Okawa, S., Saltó, C., Ravichandran, S., Yang, S., Toledo, E.M., Arenas, E., and Del Sol, A. (2018). Transcriptional synergy as an emergent property defining cell subpopulation identity enables population shift. Nat. Commun. *9*, 2595.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. Development *139*, 2288–2298.

Palii, C.G., Cheng, Q., Gillespie, M.A., Shannon, P., Mazurczyk, M., Napolitani, G., Price, N.D., Ranish, J.A., Morrissey, E., Higgs, D.R., and Brand, M. (2019). Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. Cell Stem Cell *24*, 812–820.e5.

Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A.M.-P., George, N., Fexova, S., Fonseca, N.A., Füllgrabe, A., Green, M., Huang, N., et al. (2020). Expression Atlas update: from tissues to single cells. Nucleic Acids Res. *48* (D1), D77–D83.

Papin, J.A., Mac Gabhann, F., Sauro, H.M., Nickerson, D., and Rampadarath, A. (2020). Improving reproducibility in computational biology research. PLoS Comput. Biol. *16*, e1007881.

Pei, W., Shang, F., Wang, X., Fanti, A.-K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., et al. (2020). Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. Cell Stem Cell *27*, 383–395.e8.

Peter, I.S., and Davidson, E.H. (2017). Assessing regulatory information in developmental gene regulatory networks. Proc. Natl. Acad. Sci. USA *114*, 5862–5869.

Pliner, H.A., Shendure, J., and Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. Nat. Methods *16*, 983–986.

Porter, S.N., Baker, L.C., Mittelman, D., and Porteus, M.H. (2014). Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. Genome Biol. *15*, R75.

Pratapa, A., Jalihal, A., Law, J.N., Bharadwaj, A., and Murali, T.M. (2019). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. bioRxiv. https://doi.org/10.1101/642926.

Qian, Y., Huang, H.-H., Jiménez, J.I., and Del Vecchio, D. (2017). Resource competition shapes the response of genetic circuits. ACS Synth. Biol. *6*, 1263–1272.

Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J.L., Saunders, L., Trapnell, C., and Kannan, S. (2020). Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe. Cell Syst. *10*, 265–274.e11.

Rackham, O.J.L., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., Suzuki, H., Nefzger, C.M., Daub, C.O., Shin, J.W., et al.; FANTOM Consortium (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335.

Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat. Biotechnol. *36*, 442–450.

Raredon, M.S.B., Adams, T.S., Suhail, Y., Schupp, J.C., Poli, S., Neumark, N., Leiby, K.L., Greaney, A.M., Yuan, Y., Horien, C., et al. (2019). Single-cell connectomic analysis of adult mammalian lungs. Sci. Adv. *5*, w3851.

Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467.

Roost, M.S., van Iperen, L., Ariyurek, Y., Buermans, H.P., Arindrarto, W., Devalla, H.D., Passier, R., Mummery, C.L., Carlotti, F., de Koning, E.J.P., et al. (2015). Keygenes, a tool to probe tissue differentiation using a human fetal transcriptional atlas. Stem Cell Reports *4*, 1112–1124.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. *37*, 547–554.

Sagi, I., De Pinho, J.C., Zuccaro, M.V., Atzmon, C., Golan-Lev, T., Yanuka, O., Prosser, R., Sadowy, A., Perez, G., Cabral, T., et al. (2019). Distinct imprinting signatures and biased differentiation of human androgenetic and parthenogenetic embryonic stem cells. Cell Stem Cell *25*, 419–432.e9.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. Cell *176*, 928–943.e22.

Sharma, N., Flaherty, K., Lezgiyeva, K., Wagner, D.E., Klein, A.M., and Ginty, D.D. (2020). The emergence of transcriptional identity in somatosensory neurons. Nature *577*, 392–398.

Sharpe, J. (2017). Computer modeling in developmental biology: growing today, essential tomorrow. Development *144*, 4214–4225.

Simões-Costa, M., and Bronner, M.E. (2015). Establishing neural crest identity: a gene regulatory recipe. Development *142*, 242–257.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R. (2018). Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. Cell Rep. *22*, 600–610.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., and Junker, J.P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nat. Biotechnol. *36*, 469–473.

Stumpf, P.S., Smith, R.C.G., Lenz, M., Schuppert, A., Müller, F.-J., Babtie, A., Chan, T.E., Stumpf, M.P.H., Please, C.P., Howison, S.D., et al. (2017). Stem Cell Differentiation as a Non-Markov Stochastic Process. Cell Syst. *5*, 268–282.e7.

Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322–327.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47* (D1), D607–D613.

Takasato, M., Er, P.X., Chiu, H.S., Maier, B., Baillie, G.J., Ferguson, C., Parton, R.G., Wolvetang, E.J., Roost, M.S., Chuva de Sousa Lopes, S.M., and Little, M.H. (2015). Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. Nature *526*, 564–568.

Tan, Y., and Cahan, P. (2019). SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. Cell Syst. *9*, 207–213.e2.

Teschendorff, A.E., and Enver, T. (2017). Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. Nat. Commun. *8*, 15599.

Thomson, M., Liu, S.J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875–889.

Toda, S., Blauch, L.R., Tang, S.K.Y., Morsut, L., and Lim, W.A. (2018). Programming self-organizing multicellular structures with synthetic cell-cell signaling. Science *361*, 156–162.

Toda, S., Frankel, N.W., and Lim, W.A. (2019). Engineering cell-cell communication networks: programming multicellular behaviors. Curr. Opin. Chem. Biol. *52*, 31–38.

Tonge, P.D., Corso, A.J., Monetti, C., Hussein, S.M.I., Puri, M.C., Michael, I.P., Li, M., Lee, D.-S., Mar, J.C., Cloonan, N., et al. (2014). Divergent reprogramming routes lead to alternative stem-cell states. Nature *516*, 192–197.

Tran, K.A., Pietrzak, S.J., Zaidan, N.Z., Siahpirani, A.F., McCalla, S.G., Zhou, A.S., Iyer, G., Roy, S., and Sridharan, R. (2019). Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency. Cell Rep. *27*, 1726–1741.e5.

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

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Turing, A.M. (1952). The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *237*, 37–72.

Tusi, B.K., Wolock, S.L., Weinreb, C., Hwang, Y., Hidalgo, D., Zilionis, R., Waisman, A., Huh, J.R., Klein, A.M., and Socolovsky, M. (2018). Population snapshots predict early haematopoietic and erythroid hierarchies. Nature *555*, 54–60.

Uzkudun, M., Marcon, L., and Sharpe, J. (2015). Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. Mol. Syst. Biol. *11*, 815.

van der Wijst, M., de Vries, D.H., Groot, H.E., Trynka, G., Hon, C.C., Bonder, M.J., Stegle, O., Nawijn, M.C., Idaghdour, Y., van der Harst, P., et al. (2020). The single-cell eQTLGen consortium. eLife *9*, 9.

van Hoof, D., Krijgsveld, J., and Mummery, C. (2012). Proteomic analysis of stem cell differentiation and early development. Cold Spring Harb. Perspect. Biol. *4*, 4.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. Nature *563*, 347–353.

Wagner, F., and Yanai, I. (2018). Moana: A robust and scalable cell type classification framework for single-cell RNA-Seq data. bioRxiv. https://doi.org/10.1101/456129.

Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science *360*, 981–987.

Wang, Y.X.R., and Huang, H. (2014). Review on statistical methods for gene network reconstruction using expression data. J. Theor. Biol. *362*, 53–61.

Weinreb, C., and Klein, A.M. (2020). Lineage reconstruction from clonal correlations. Proc. Natl. Acad. Sci. USA *117*, 17041–17048.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, 367.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018.

Yachie-Kinoshita, A., Onishi, K., Ostblom, J., Langley, M.A., Posfai, E., Rossant, J., and Zandstra, P.W. (2018). Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. Mol. Syst. Biol. *14*, e7952.

Yang, J., Ryan, D.J., Lan, G., Zou, X., and Liu, P. (2019a). In vitro establishment of expanded-potential stem cells from mouse pre-implantation embryos or embryonic stem cells. Nat. Protoc. *14*, 350–378.

Yang, P., Humphrey, S.J., Cinghu, S., Pathania, R., Oldfield, A.J., Kumar, D., Perera, D., Yang, J.Y.H., James, D.E., Mann, M., and Jothi, R. (2019b). Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. Cell Syst. *8*, 427–445.e10.

Yao, Z., Mich, J.K., Ku, S., Menon, V., Krostag, A.-R., Martinez, R.A., Furchtgott, L., Mulholland, H., Bort, S., Fuqua, M.A., et al. (2017). A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development. Cell Stem Cell *20*, 120–134.

Yordanov, B., Dunn, S.-J., Kugler, H., Smith, A., Martello, G., and Emmott, S. (2016). A Method to Identify and Analyze Biological Programs through Automated Reasoning. NPJ Syst. Biol. Appl. *2*, 2.

Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. Hum. Mol. Genet. *24* (R1), R102–R110.

Zhang, A.W., O'Flanagan, C., Chavez, E.A., Lim, J.L.P., Ceglia, N., McPherson, A., Wiens, M., Walters, P., Chan, T., Hewitson, B., et al. (2019). Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. Nat. Methods *16*, 1007–1015.

Zhang, K., Wu, D.-Y., Zheng, H., Wang, Y., Sun, Q.-R., Liu, X., Wang, L.-Y., Xiong, W.-J., Wang, Q., Rhodes, J.D.P., et al. (2020). Analysis of Genome Architecture during SCNT Reveals a Role of Cohesin in Impeding Minor ZGA. Mol. Cell *79*, 234–250.e9.

Zhou, Q., Chipperfield, H., Melton, D.A., and Wong, W.H. (2007). A gene regulatory network in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *104*, 16438–16443.

Zhou, J.X., Brusch, L., and Huang, S. (2011). Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model. PLoS ONE *6*, e14752.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell *65*, 631–643.e4.