Computational and Structural Biotechnology Journal 18 (2020) 2733–2743

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Probing pluripotency gene regulatory networks with quantitative live cell imaging

Anne L. Plant *, Michael Halter, Jeffrey Stinson

Biosystems and Biomaterials Division, National Institute of Standards and Technology, United States

ARTICLE INFO

Article history:
Received 1 July 2020
Received in revised form 14 September 2020
Accepted 15 September 2020
Available online 20 September 2020

Keywords:
Systems Biology
Quantitative light microscopy
Live cell imaging
Phenotype heterogeneity
Dynamic cell response
Dynamic fluctuations
Fluorescent reporter cell lines
iPS cells
Transcription factors
Pluripotency

ABSTRACT

Live cell imaging uniquely enables the measurement of dynamic events in single cells, but it has not been used often in the study of gene regulatory networks. Network components can be examined in relation to one another by quantitative live cell imaging of fluorescent protein reporter cell lines that simultaneously report on more than one network component. A series of dual-reporter cell lines would allow different combinations of network components to be examined in individual cells. Dynamical information about interacting network components in individual cells is critical to predictive modeling of gene regulatory networks, and such information is not accessible through omics and other end point techniques. Achieving this requires that gene-edited cell lines are appropriately designed and adequately characterized to assure the validity of the biological conclusions derived from the expression of the reporters. In this brief review we discuss what is known about the importance of dynamics to network modeling and review some recent advances in optical microscopy methods and image analysis approaches that are making the use of quantitative live cell imaging for network analysis possible. We also discuss how strategies for genetic engineering of reporter cell lines can influence the biological relevance of the data.

Contents

* Corresponding author.
E-mail address: anne.plant@nist.gov (A.L. Plant).

https://doi.org/10.1016/j.csbj.2020.09.025

1. Introduction

Quantitative live cell imaging contributes unique data and insights to the study of pluripotency, but gene regulatory networks (GRN) are most often examined by measurement of gene transcripts and protein–protein interactions (PPI) as reviewed in [1]. Whether these measurements are made on a population of cells or are determined on the level of individual cells, they are endpoint measurements in that the cells are sacrificed in the process. Temporal response data has to be inferred from measurements of different populations of cells. Transcriptomics and PPI data have identified large numbers of putative network components and have provided classifications of gene expression products. Single cell transcription studies have illuminated, for example, the gene expression profile changes that accompany hematopoiesis [2]. Other studies have provided information about the heterogeneity of gene expression in individual cells within an isogenic population and insight into the relative degree of variability between individual cells at early stages versus late stages of reprogramming to iPSC (for example, see [3]).

Data from static methods such as transcriptomics and flow cytometry can indicate correlations in the appearance of different network components in populations of cells over time or in response to perturbations and can suggest causation between gene products. However, it is clear that causality is impossible to confirm from single point-in-time studies [4,5]. While the value of omics measurements is great, dynamic data of expression of network components in individual cells in real time is critical for determining unambiguously how networks control cell fate [4–7]. Details regarding the controlling features of many networks are still under investigation. Many network variables have been identified as contributing to the control of pluripotency [8]. However, there is still significant ambiguity even in the relationship between the canonical factors OCT4, SOX2 and NANOG [9]. This makes it challenging to unambiguously characterize the state of the population. Furthermore, knowing the dynamic relationships between network components makes it possible to predict how rapidly the state of the population will change under changing conditions, such as was demonstrated in [10]. Fig. 1 summarizes the observations that have been reported for the relationship between the canonical transcription factors and pluripotency substates.

In contrast to measurements of transcripts and other single time endpoint measurements like flow cytometry, live cell imaging uniquely allows individual cells to be probed sequentially over time. Quantitative imaging has been used to directly measure time series gene expression dynamics in individual cells [7,11–15]. In these studies, a fluorescent protein (FP) is expressed at the same time as a gene of interest (GOI) and reports on the activation of that gene. The term 'image-based systems biology' has been used to refer to the use of imaging to take advantage of spatial and temporal information and link theoretical and experimental analysis of biological processes [16]. Phenomenological observations of populations of individual cells in time have provided insight into mechanistic details that cannot be directly detected from measurements of single cells at a single point in time. For example, Maherali [17] used green fluorescent protein (GFP) as a NANOG reporter to show that pluripotency could be conferred to individual somatic cells by fusing them with iPS cells, and then using another GFP reporter, they demonstrated the inactivation and reactivation of the X-chromosome in individual cells upon induction of a pluripotent state. Live cell imaging has allowed direct observation of cell lineage progression and cell fate [18]. Live cell imaging has helped to elucidate other information about stem cells that would not be accessible from methods that require cell destruction. This includes the direct measurement of individual cells in the context of surrounding cells, such as the preferential location of self-renewing embryonic stem cells in the interior of a colony [19], the analysis of correlations of gene expression with other directly measured characteristics such as division time and symmetry of division [18]; and the determination of lineage decision points in neural stem cells [20]. By following individual cells over time, Strasser et al. [21] used correlations in marker expression in related progeny to determine the time delay in expression following lineage commitment in hematopoiesis. Live cell imaging provides unique data that can verify direct correlations of gene expression within individual cells with other features of that cell's behavior.

Gene regulatory networks associated with pluripotency have been studied directly with live cell imaging through analysis of the dynamics of expression of transcription factors associated with FPs [7,15,21,22]. Time-resolved measurements in individual cells have allowed assessment of the dynamics of NANOG fluctuations

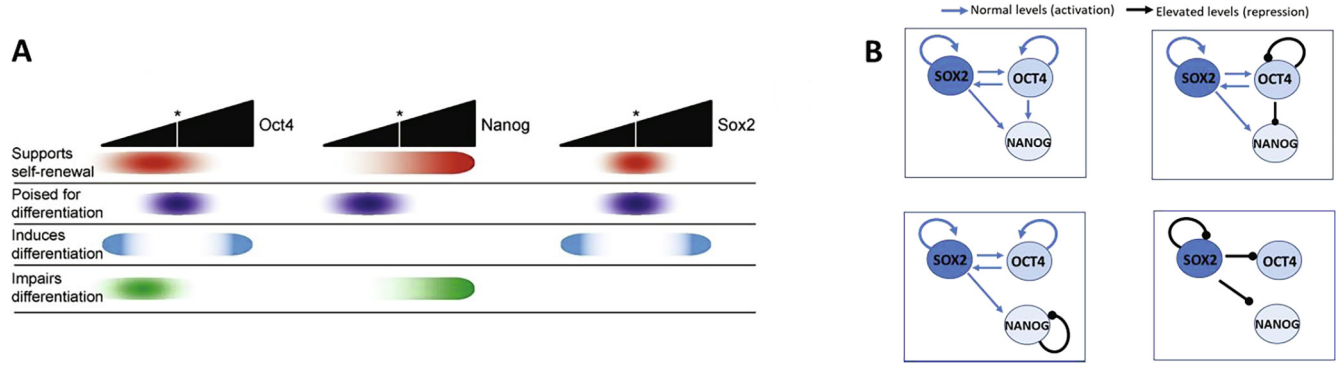

Fig. 1. A. Published studies reviewed in [9] indicate an ambiguity in TF levels and their relationship to possible substates of pluripotency, suggesting that pluripotency includes substates that are highly stable and substates that are more easily differentiated. B. The proposed interactions between OCT4, SOX2 and NANOG include auto- and cross-regulation of related TFs by positive and negative feedback.

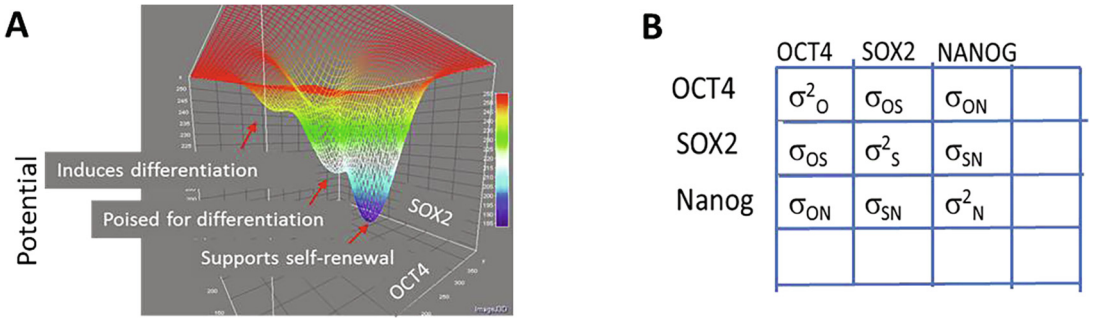

Fig. 2. A. A conceptual visualization showing a possible two-dimensional landscape demonstrating the relationship between OCT4 and SOX2 over a population of pluripotent cells. The expression of two GOI can be monitored simultaneously in individual cells by engineering dual reporter cell lines. Combining data from a series of dual reporter cell lines would allow a multidimensional landscape of network component interactions to be constructed. B. Network variables can be examined in a pairwise fashion using a series of dual-reporter engineered cell lines. A 3 × 3 matrix of TFs is shown here, but there is, in principle, no limit on the number of transcription factors that can be examined with appropriately engineered cell lines.

and the development of models that predicted the steady state distribution of NANOG expression levels [22]. Quantifying fluctuations in the expression of a single gene in individual cells over time directly by live cell microscopy has been reported [7,10,22]. However, although dual-reporter cell lines have been engineered [7,23–27], measuring trajectories of expression of more than one gene in individual cells has not been reported to our knowledge. Correlations in expression of multiple gene products can indicate mutual control of those genes by upstream factors [28], and correlated responses between network components provides a higher degree of confidence in the nature and strength of the relationship between gene products such as transcription factors [29] than can be achieved by static measurements such as flow cytometry or transcriptomics. By engineering a series of multi-reporter cell lines, and with appropriate imaging technologies, the measurement of correlations in the expression of multiple gene products is possible. By quantifying coordinated fluctuations in pairs of transcription factors in individual cells within a series of engineered cell lines, one can, in principle, establish an N dimensional diffusion matrix, where N is the number of gene products being examined (Fig. 2). The rate constants for these dynamic correlations would allow predictions of the time required for the population to respond to a perturbation, as was shown for a one-dimensional system [10].

A critical advantage of imaging is that the dynamic relationships between network components may demonstrate relationships to other phenotypic characteristics of the cells, such as relative location in a colony, morphological characteristics, proximity to other cells, division time, metabolic state, etc. These features in combination add knowledge that cannot be achieved from data from a single point in time.

2. Advances in imaging technologies for probing multidimensional GRN dynamics

2.1. Quantitative live cell imaging provides access to patterns of gene expression and gene regulatory dynamics

Despite the temporal and spatial information that live cell imaging uniquely offers to network analysis, it is not employed as frequently as end point measurement methods. There are challenges associated with quantitative live cell imaging, and advances in overcoming them are enabling the wider-spread use of imaging in systems biology studies. One challenge is collecting accurate fluorescence intensity measurements from time lapse images of FP reporter cells. There are many factors that can influence the bias and uncertainty of microscopy-based fluorescence measurements, including non-uniform illumination, background fluorescence, and image signal to noise ratio, as recently reviewed in detail [30]. Long-duration time lapse imaging and quantitative analysis present several additional challenges including maintaining constant environmental conditions, avoiding photobleaching and phototoxicity, and accurate segmentation and tracking [31]. Innovations in illumination sources, detectors, filters, and other hardware are improving the feasibility of quantitative long-term kinetic studies on live cells [32,33]. More recent techniques such as light sheet microscopy have benefited from many of these advances, as have methods such optical ptychography and continuous motion imaging which are in the early stages of application to live cell microscopy.

Another challenge is that network analysis requires the acquisition of large volumes of image data. As an example, as shown in Fig. 3, quantitative live cell imaging was used to measure the fluctuations in expression of a single gene [10]. Individual cells were segmented and tracked over time and the change in GFP intensity in each cell was recorded every 15 min. Trajectories were then analyzed by mean square displacement to determine a fluctuation rate constant. This study involved the collection of 36 field of views at 15 min time intervals. The area of each field of view was approximately 1 mm^2, so a 6 mm^2 area was probed at each timepoint. Overall, the study contained approximately 23,000 single cell intensity measurements in time series. To perform a similar analysis for a higher dimensional landscape that considers multiple network components, much more data will be required; the analysis of the interactions between d number of genes will require the number of time series samples in individual cells raised to the power of d. Using the example in [10] as a guide, an analysis interrogating two genes (d = 2) would require a 36 mm^2 area (or equivalently 1296 fields of view) to be imaged at each time point to reach a similar level of sampling. This rate of imaging is not possible with current "stop and stare" microscopy [34], but it can be achieved with methods such as those discussed below.

While the rapid rate of imaging that can be achieved presents a significant data challenge, the availability of software tools for managing and analyzing these larger datasets is becoming increasingly available [37–39]. Eliceiri et al. [40] provide a comprehensive overview of all the different software tools that are necessary for the implementation of any quantitative imaging workflow. A more recent review evaluates 25 different open source software platforms for the analysis of cell image data [41]. These general-purpose tools facilitate the building of quantitative workflows for the analysis of live cell imaging data. Software tools designed specifically for the analysis time lapse image data include LEVER [42], Lineage Mapper [36], Sequitr [43] and The Tracking Tool, tTt [38]. Together, these tools are facilitating the quantitation of cellular dynamics by time lapse microscopy.

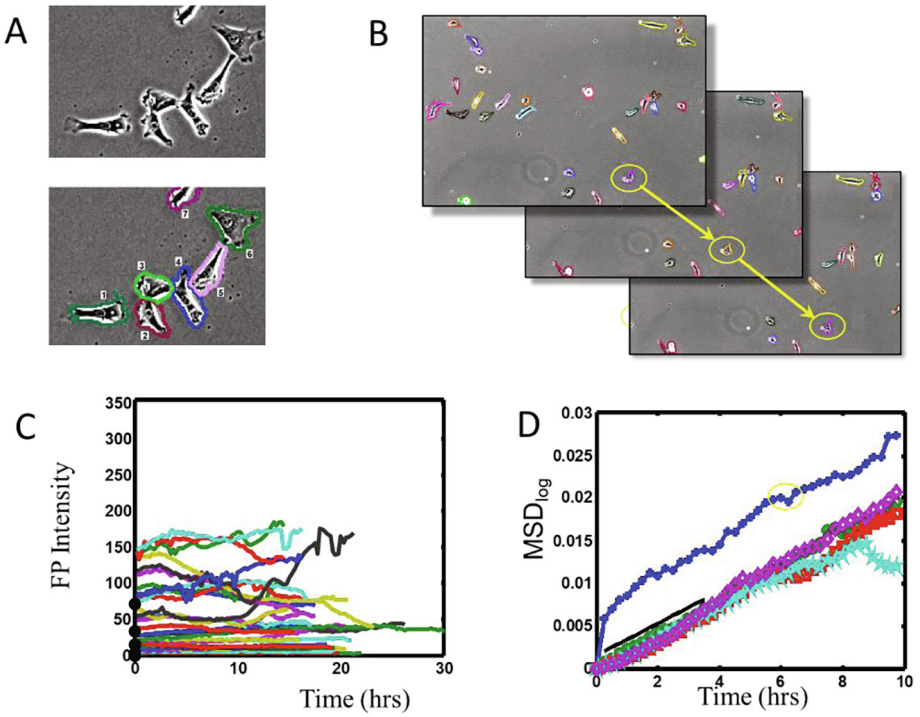

Fig. 3. Example workflow for acquiring and analyzing time series data. A. Image analysis is performed to identify and segment individual cells from transmitted light images [35]. B. Individual cells are tracked between successive images in a time lapse series [36]. Intracellular fluorescence intensity is computed using the segmentation mask from the transmitted light image applied to the corresponding fluorescence image. C. Time series fluorescence intensity data are plotted for numerous single cells from a single experiment [10,12]. D. The single cell time series intensity data in C are analyzed by a mean square displacement analysis to quantify the rate of fluctuations in log (FP) expression [10].

2.2. Light sheet microscopy for low perturbing, rapid fluorescence data acquisition

Light sheet microscopy has been applied most often to organoids and other 3-dimensional spatial arrangements of cells. The technique provides advantages to cells in monolayer culture as well. The advantages of light sheet microscopy are primarily based on the rapid scanning mode of the technique. Rapid scanning has a number of advantages to network analysis: it allows for data to be accumulated in a time frame that is consistent with gene expression fluctuations, and individual cells are exposed to excitation light for short periods of time which reduces the occurrence of phototoxicity. In addition, the background signal from media surrounding cells is minimized, thus improving the signal to noise ratio. Several implementations of light sheet microscopy have been already been described that are suitable for live cell imaging of pluripotent stem cells in multi-well culture plates [44,45]. In addition, super resolution lattice light sheet imaging has been applied to the analysis of SOX2 binding to heterochromatin regions in mouse embryonic stem cells [46].

2.3. Large spatial bandwidth imaging can be used to monitor dynamics in large numbers of single cells

Compared to illumination for fluorescence excitation, transmitted light imaging enables larger amounts of image data to be collected under less perturbing conditions. Fundamental changes to traditional imaging strategies for time lapse imaging are taking place in the form of improved spatial bandwidth, i.e., the relationship between the temporal resolution of a microscope and the area that can be imaged within the timeframe. Imaging with a small time interval between frames can improve cell tracking accuracy, but with traditional imaging methods this would come at the expense of the number of fields of view that could be sampled in that time frame, and therefore the numbers of cells that could be imaged. High spatial bandwidth product imaging techniques can allow for time lapse data to be acquired from a large area with high temporal frequency, increasing the efficiency of experimental data acquisition.

Two examples of high spatial bandwidth product methods are continuous motion imaging and optical ptychography. The approach for continuous motion imaging was first implemented for optical microscopy using specially designed time delay integration cameras [47]. Continuous motion imaging has become standard practice for whole-slide imaging devices for pathology yielding high throughput acquisition (~15 mm^2/min) of imaged samples. The approach of continuous motion imaging has recently been applied successfully to live cell imaging studies [48]. Transmitted light microscopy imaging of cell in culture vessels allowed scanning of the sample at much higher speed (up to 30 times faster) than conventional methods. The cellular sample under investigation moves continuously and is captured using a flash illumination which creates an exposure time short enough to prevent motion blur.

Optical ptychography is another approach that can increase the acquisition rate of dynamic cellular data by microscopy. Optical ptychography achieves both a large field of view and high spatial resolution through the application of computational phase reconstruction using multiple images as inputs, each image resulting from a different illumination configuration [49]. In one embodiment, an LED array is the transmitted light source that illuminates the specimen during multiple camera exposures, each using a different subset of LEDs in the array [50]. Numerous recent advances in hardware and software have been reported for optical ptychography implementations. The cell biology applications of optical ptychography remain limited, most of which focus on the label-free, high contrast and quantitative phase aspects of optical ptychography [51]. These applications demonstrate the feasibility of incorporating optical ptychography into a quantitative cytometry

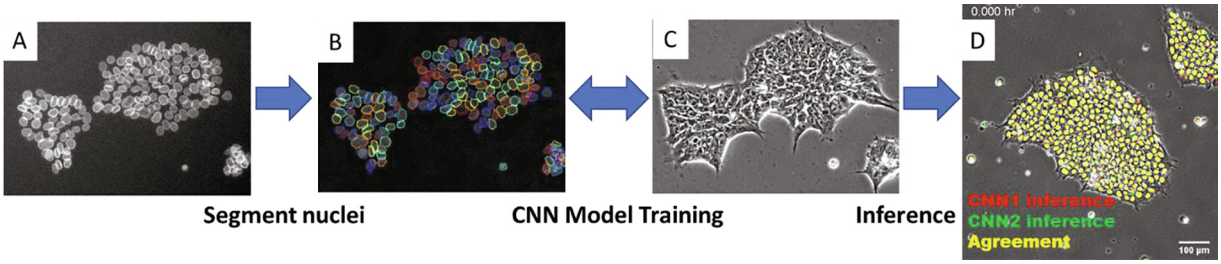

Fig. 4. Application of a CNN [42] to predict nuclear segmentation from phase contrast images. A. This example uses a cell line containing a fluorescent nuclear marker (WTC-mEGFP-LMNB1-cl210 from the Allen Institute for Cell Science). B. The presence of fluorescence enables accurate nuclear segmentation because of the high signal to noise ratio. C. The segmented images are used to train a CNN model that uses transmitted light (phase contrast) images as input. D. Multiple models were trained using different subsets of training data. The observed agreement between the two models indicates a degree of robustness for training CNNs to segment iPSC nuclei from phase contrast images.

workflow. Looking to the future, the high spatial bandwidth product capacity of optical ptychography also means it is an appealing technique for acquiring dynamic data fundamental to GRNs.

Both continuous scanning microscopy and optical ptychography speed the acquisition of transmitted light imaging and will enable the tracking of larger numbers of cells even as they move, change shape and divide. Strategies for integrating fluorescence imaging within the acquisition protocol will make both tracking of large numbers of cells and quantitation of fluorescence signal in each cell at appropriate timeframes possible [52].

2.4. Convolutional neural networks aid post-image acquisition processing and data workflows

A recent advance that is enabling imaging as a quantitative cytometry tool is the application of image analysis pipelines based on artificial intelligence (AI) computational methods [53], and particularly convolutional neural networks (CNNs). The number of implementations of CNNs on optical microscopy data has exploded. The U-Net architecture, for example was published in 2015 [54] and has over 14,000 citations to date. Compared to feature-based algorithms, CNNs are powerful tools for probing gene regulatory networks with optical microscopy. CNNs allow the use of high signal to noise fluorescence data of single-time point images to be used for training and testing pipelines that can then be deployed on label free live cell images [55,56]. An example of this application is shown in Fig. 4. The ability to image and track individual cells using transmitted light without fluorescence excitation reduces phototoxicity, enables rapid data acquisition, and increases the availability of fluorescence channels to be used for monitoring gene expression dynamics. Furthermore, the development of image analysis algorithms using AI are proving to be less labor-intensive than feature-based algorithms [57] and so are allowing quantitative live cell imaging studies to be conducted that were not previously possible. Of particular relevance to studying GRNs, CNNs can be retrained to accommodate changes in cell features.

2.5. Outlook

Hardware and software advances such as those discussed above will allow image data to be acquired with sufficiently high temporal resolution to achieve high accuracy tracking of individual cells over long times while minimizing cellular damage. With sufficient temporal and spatial resolution, dynamical data of gene expression can be correlated with other dynamical data such as migration rate, rate of cell division, rate of divergence of expression between cells after division, etc. Data on statistically relevant numbers of cells can be collected making it possible to use quantitative live cell imaging as a systems biology tool to more completely inform our understanding of GRNs.

3. Examples of mathematical modeling of dynamics in networks

3.1. The role of dynamic fluctuations in networks

Time-series measurements of individual cells provide dynamic data that inform models and predict observations of cell population responses. Dynamic fluctuations in gene expression are essential for maintaining the steady state heterogeneity in phenotype that is observed in an isogenic population of cells. A source of fluctuations, transcriptional bursting, has been documented with single molecule studies and subsequent time-lapse single cell imaging measurements which have provided direct evidence of the effect of this type of "noise" on population heterogeneity [58]. It is now recognized that the observed heterogeneity in cell populations is due to fluctuations in biochemical reactions in general, and encodes mechanistic information about intersecting signaling pathways, complex regulatory mechanisms, and signaling pathway promiscuity [59–62]. Dynamic fluctuations are therefore a critical component of regulatory networks. Dynamic data of phenotype from individual cells provides access to the net effect of all molecular contributions to that phenotype. Experimental imaging studies [10,13,15,22,63,64] and theoretical studies [65–69] indicate that the dynamic fluctuations that result in the observed heterogeneity of phenotypes play a critical role in directing cell response, and in maintaining homeostasis [29].

An example of the prominent role that short-time fluctuations in NANOG expression levels can play in influencing the outcome of a regulatory network of transcription factors was demonstrated in a theoretical study [65]. Using a master equation approach and the Gillespie algorithm [70], the authors inferred the role that fluctuations in NANOG levels in cells might have on cells transitioning between pluripotent microstates and between pluripotency and differentiation. With differential equations they modeled a simplified network of NANOG, OCT4-SOX2 complex, fibroblast growth factor (FGF), and a differentiation gene. They assumed that at steady state each network component fluctuates around its mean value with a Gaussian distribution. By making assumptions about the effect of NANOG levels on differentiation, they showed that when low levels of NANOG correlated with increased levels of FGF2 or differentiation factor, this would lead to downregulation of OCT4-SOX2 complex and transition to differentiation. External factors (2i/3i media) played a key role in network outcome by reducing fluctuations in NANOG expression.

3.2. Potential energy landscapes

The range of phenotypes expressed by a cell population is a probability density function, reflecting the probability, $P(x)$, with which each microstate, x, occurs within the population. Every cell

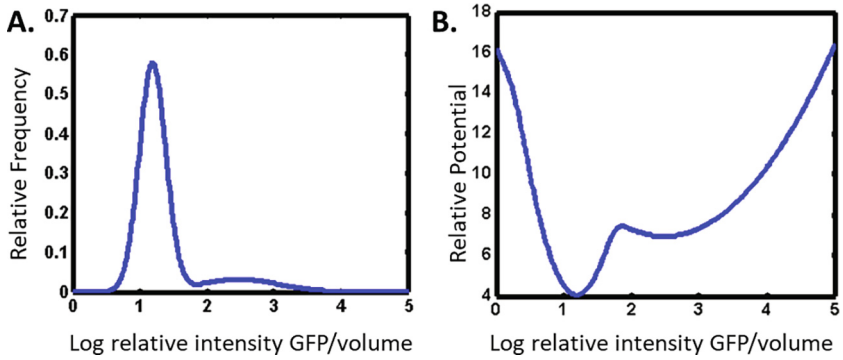

Fig. 5. The relationship between a population distribution and potential landscape. A. An example of a steady state distribution, P_{ss}, of gene promoter activity across the population plotted as the log of reporter FP intensity. This plot represents the heterogeneity observed across the population due to differences in expression levels by individual cells. If perturbed, the population will return to this steady state distribution. B. A corresponding potential landscape is determined as $U = -\ln(P_{ss})$ and indicates areas of relative low potential energy and an energy barrier between them. These plots take into account the confounding effect of increased proliferation of high expressing cells that was observed in this study [10].

in the population occupies a microstate within the set of possible microstates that exist for that population under those conditions. All cells have the potential of occupying any of the possible microstates, but some microstates have a higher probability of being occupied than others. At the steady state the distribution of phenotypic states, P_{ss}, can be represented as a potential landscape by applying the function $-\ln(P_{ss})$, as shown in Fig. 5, in which the probability of microstate occupation is indicative of the thermodynamic stability of that microstate. Landscapes are a useful way to present data from live cell imaging, flow cytometry, and single cell transcriptomics analysis, and are used in different kinds of modeling approaches. Landscape models guide thinking about the dynamics of transitions that can occur in pluripotency and other kinds of networks, beginning with the often-cited work of Waddington and including many others [10,59,71–77].

Probabilistic fluctuations in gene expression explain how cells that are isolated from a population based on their instantaneous phenotype will produce progeny that will eventually recapitulate the original distribution of phenotypes [10,22,78,79]. For example, Chang et al. [78] used flow cytometry to study the heterogeneity of the stem cell marker Sca-1 across a population of stem cells. They sorted the population into subpopulations with different Sca-1 expression levels. The subpopulations demonstrated different proclivities for differentiation into either the erythroid or myeloid lineage. Each had distinct transcriptomes but relaxed back to the original population distribution over a period of time that allowed 12 population doublings. The kinetics of population relaxation could not be explained by a simple process of adding noise to a deterministic equilibrium state but required invoking a complex landscape with multiple quasi-stable states. In a study from Kalmar et al. [22], a population of pluripotent cells expressing fluorescent reporters for NANOG demonstrated a bimodal distribution of NANOG levels by flow cytometry; selecting and culturing a subpopulation of cells resulted in recapitulation of the original distribution. The data were modeled using differential equations and results showed that fluctuations in NANOG levels were essential for the role that NANOG seems to play as a determinant of differentiation.

3.3. Direct determination of the kinetics of fluctuations in single cells

Live cell imaging provides the opportunity to directly measure the rates of fluctuation in a gene of interest [10,15]. Employing both quantitative live cell imaging and flow cytometry, Sisan et al. [10] observed cells isolated from a population that produced green fluorescent protein (GFP) driven by the promoter for the extracellular matrix protein tenascin-C. Four subpopulations with distinct GFP intensities were allowed to relax back to the steady state distribution over long times. In this study, the rate constant for fluctuations in expression of the tenascin-C gene, determined as shown in Fig. 3, allowed excellent prediction of the complex kinetics of relaxation. The analysis demonstrated that the kinetics with which an individual cell can recapitulate the stationary population distribution is determined by the rate of fluctuation in gene expression and its position in the landscape.

The analysis used by Sisan et al. [10] was a Langevin/Fokker-Planck approach. This is a coarse-grained approach in which the Langevin equation identifies two predominant features of the system. One feature is a deterministic component, a force, which corresponds to the landscape shape which is derived from the measurement of the distribution of expression levels across the population of cells. The second feature, the diffusion coefficient, is the rate of fluctuation in gene expression and is measured directly in the cells as a mean square displacement in intensity of the FP probe over time.

This coarse-grained approach requires only data that is experimentally measurable, i.e., the distribution of individual cell responses across the population, and the measured mean square displacement of single cell intensities over time. In contrast, modeling with differential equations requires assumptions about rates and binding constants, which are often poorly known, and is computationally more expensive with increasing network size. The Langevin equation approach, which provided a numerical solution through simulation, allowed an excellent prediction of the 4 different nonlinear relaxation rates for 4 subpopulations of cells that were isolated by flow sorting.

3.4. Correlations in fluctuations can indicate network organization and strength of interactions between network components

Fluctuation rates in expression of fluorescent reporters have been measured directly with live cell imaging of FP-expressing fibroblasts [10] and embryonic stem cells [15]. Live cell imaging in principle allows simultaneous examination of multiple network components in individual cells and quantification of the dynamic relationships between those components. Theoretical studies have addressed the importance of quantifying the extent to which fluctuations in multiple network components are correlated as a true determinant of causality. Frequently causality is inferred from correlated occurrence (for example in clustering analysis from tran-

scriptomics measurements), but correlation does not guarantee causation [4,5]. Theoretical studies have demonstrated that correlated dynamic fluctuations indicate causal influence between multiple components of regulatory networks [4,29,68].

Hubbard et al. [29] showed that the magnitude of correlation in fluctuations between two or more putative network components, together with the landscape gradient of their expression levels, has a thermodynamic meaning that is related to the entropic price of the organized network structure, and to the amount of dissipative heat associated with supporting the operation of the homeostatic network. An example of measuring correlations in fluctuations between 2 network components was demonstrated in a synthetic gene regulatory circuit in bacteria [63]. So far, a direct measure of correlations in the fluctuations in multiple transcription factors in mammalian cells has not been reported to our knowledge. However, dual reporter cell lines that might be used for such an analysis where FPs are associated with two different genes associated with pluripotency and differentiation networks have been reported [7,23,24,26,27,80]. Fig. 2 depicts a two-dimensional landscape of OCT4 and SOX2 expression that might be constructed from measurements of a population of a single dual reporter pluripotent cell line. A higher-dimensional landscape would result from a series of cell lines reporting on different pairs of transcription factors.

The power of an analysis of correlated fluctuations within a network will be great, as it has the potential to identify the strength of causal relationships between network components. The promise of such an analysis includes a thermodynamical understanding of network structure, and identification of the most important contributors to a regulated network [29].

3.5. Comparing and combining omics and imaging data

Live cell imaging allows individual cells to be followed over time continuously, enabling direct observation of the progression of multiple intracellular biochemical events. Such direct observations can strengthen our quantitative understanding of the relationship between one molecular event in time and another event in time, and the variability in those processes and relationships over the population. This additional information enables predictive models. Omics methods can provide data on thousands of molecular species in a population of cells, and single cell RNAseq typically provide data on dozens of gene transcripts in each cell. However, endpoint measurements fail to capture dynamics in network operation and cannot provide evidence of causality between network components [5]. Imaging is best suited for in- depth kinetic studies of a relatively small subset of components of a putative GRN. Clearly a combination of both methods will be ideal to fully understand and predict the activity of GRNs.

4. Appropriate engineering and characterization of reporter cell lines for network analysis

4.1. The importance of fluorescent protein reporter cell lines for studying networks

FP-producing cell lines that report on genes involved in pluripotency and differentiation have enabled the understanding of these complex regulatory systems. Numerous pluripotent cell lines designed with fluorescent reporters of gene expression have been created [24]. One of the earliest human reporter cell lines was from the work of Zwaka and Thompson [81]. They adapted a mouse electroporation protocol for use with human ES cells, taking into account the larger size of human stem cells. Cells were edited by spontaneous homologous recombination to insert a copy of the gene for GFP downstream of OCT4 with an internal ribosome entry site (IRES) translational read-through signal.

The advent of genomic editing aided by TALENs [82], Zn fingers [83], and CRISPR technologies [84] have enhanced the design of reporter cell lines by allowing greater efficiency in numbers of cells edited, and greater targeted sequence specificity of edits in the genome. In the earlier work of Zwaka and Thomson [81] that used spontaneous homologous recombination, an approximate recombination rate of 1 in 10^5 cells was reported. Current methods incorporating genomic editing aided by TALENs, Zn fingers or CRISPR has greatly improved specific gene targeting efficiencies by several orders of magnitude [85]. These approaches have made it practical to engineer stem cell lines with single [14,86–92], and dual [7,24,26,27,93] gene reporter systems. Dual gene reporters have also been generated in mouse cells through crossing of mice containing a single gene engineered reporter [23,25,80].

4.2. Design considerations associated with genetic modification

Assuring the biological relevance of cell lines and FP probes is critical if they are to be useful tools for elucidating GRNs. There are important caveats associated with the options for designing reporter cell lines for the purpose of studying pluripotency and GRNs. Currently, design of knock-in genome edited fluorescent reporter cell lines have been generated using the strategies listed in Table 1.

Each of the designs listed in Table 1 uniquely impacts the ability of the FP expression to act as a biologically relevant surrogate signal for the expression of the GOI. The first example (A), assumes that a defined upstream region (usually 1 k–2 k bases) is an adequate replacement for the complex regulatory domain of the GOI's endogenous chromosomal location. Eukaryotic genes are often regulated by complex transcriptional control elements which can include both positive and negative elements far upstream, downstream, and/or within introns. Replacing a copy of the GOI with a FP coding sequence within the endogenous chromosomal milieu, as described in (B) eliminates this problem, but introduces a change in the allelic dosage of the GOI. Smith et al. [28] developed a mathematical model to predict the effects of particular FP reporter expression designs on the ability of an engineered cell line to accurately predict wild-type transcriptional activity which they termed "reporter accuracy". Using this model, they compared wild type (WT), and the methods of B, C and D above.

Their model predicted that scenario B would introduce the greatest perturbation of transcriptional response from WT, with scenarios c and d having less of an impact. To test against the predictions of the model, they used a cell line containing a FP "knock-in" replacing NANOG gene, as in scenario B and compared it with

Table 1
Editing strategies used in creating FP reporter cell lines.

	Editing Strategies for FP Reporter of GOIs	References
A	The FP coding sequence is inserted downstream of a putative promoter sequence at a "safe harbor site"	[94]
B	The FP coding sequence is inserted downstream of endogenous regulatory elements, replacing the GOI	[95]
C	The FP coding sequence immediately follows or precedes the GOI to create a single transcript and results in a fusion protein with the GOI at N- or C-terminal, usually with peptide linker	[14,28,87,90,96]
D	The FP coding sequence is inserted downstream of endogenous regulatory elements, and separated from it with a translational read-through element such as 2A or IRES so that a single transcript is ultimately translated as separate protein units analogous to a bacterial operon	[25,89,91]

the WT cell line without the FP "knock-in". The results supported the model indicating that altering the allelic dosage of NANOG through the "knock-in" introduced a significant change from the wild type expression levels that would confound data interpretation of biological effects. Other work [25] using a dual NANOG fluorescent reporter cell line created by sequential gene targeting (one allele with the FP EGFP- and the second with the FP mCherry and a read-through design) supports the critical role gene dosage plays in the regulation of NANOG and pluripotency. It is reasonable to hypothesize that allelic dosage generally is critical to transcriptional control in pluripotency because of the unique chromosomal architecture of stem cells [97,98]. The use of fusion proteins (C, above) is common and has the advantage of providing co-localization information with the GOI. However, careful evaluation of the potential impact of the fused FP-GOI complex must be considered. Confirmation that characteristics such as turnover rates, localization and functionality are necessary. Especially if the GOI is a transcription factor (TF), the FP fusion may cause unpredictable effects on TF functionality that are difficult to detect because of the complex interactions of TFs with other GRN components.

Reporter designs utilizing translational read-through elements for the FP expression under the endogenous transcriptional control of the GOI (D, above) limit the risks associated with those previously described. The transcription of the two components are linked genetically, but the translational products are produced independently. There is no alteration of the gene dosage and minimal disruption of the normal chromosomal gene regulatory environment. Although it is possible that the introduction of additional DNA upstream or downstream of the GOI could alter the transcriptional regulation, that risk appears relatively slight in comparison, especially for downstream insertions. It should be noted that abundance of the FP cannot be used as an accurate reporter of the TF concentration or the lifetime of the GOI since the protein degradation rates are unlikely to be identical, and the FP will exhibit a lag between translation and fluorescence because of the maturation time required for fully functional folding. However, because the transcripts are co-expressed, the fluctuations in the rate at which the transcriptional activation of the GOI will be well approximated.

Another design element to consider is the addition of a nuclear localization sequence by fusion to the FP [99]. This can be advantageous for two reasons. The soluble protein will be concentrated in the nucleus, resulting in an isolated region of interest from which to quantify the fluorescence and providing a higher number of counts over a smaller number of pixels which will increase the signal to noise ratio. In addition, in many cases the nucleus is easier to segment and distinguish as a unique object than is the cell body, resulting in a more accurate tracking of live cells.

4.3. The importance of characterization and reporting of details for edited cell lines

A number of reporter cell lines are commercially available or published on, yet very few seem to be shared and redeployed among the researcher community, even though considerable time and resources are required to generate new fluorescent reporter cell lines. Incomplete characterization of cell lines and insufficient reporting of details for creating and validating cell lines is common and may be an impediment to their adoption by others. While cell line characterization is quite extensive in some reports [14,17,26,96] too few are sufficiently documented. This failure leads to a lack of confidence in the biological significance of the data generated using these cell lines. Exemplars of thorough cell line validation and characterization include the Allen Institute for Cell Science (AICS) [96] and the European Molecular Biology Laboratory (EMBL) [87]. Both groups focus on the complete description of the reporter construction, the use of multiple methodologies to interrogate the details of the genomic editing accuracy and thorough evaluation of the functional properties of the reporter that relate to its ability to act as a biologically relevant expression surrogate for the GOI. This level of qualifying cell lines and reporting those data provides confidence in the biological significance of research data generated using the cells lines. Clearly, what is necessary to increase confidence in reporter cell lines is the inclusion of detailed protocols for all aspects of the reporter cell line development process. Critical information provided by AICS and EMBL include genome edit design and methodology, copy number analyses (digital droplet PCR and/or Southern blot), definitive details on the insertion event (for example with junctional PCR and Sanger sequencing of insertional context), evaluation of the presence of unintended changes to the genome, and confirming normal biological function of the edited line (including, for example, protein expression quantification, cell cycle timing, pluripotency marker verification, and differentiation potential).

4.4. The complications of pseudogenes for genome editing and the need for effective screening strategies

Another factor that is rarely, if ever, considered in designing and validating reporter FP cell lines is the potential complication represented by pseudogenes. A survey of genes tightly associated with pluripotency and oncogenesis reveals that a majority of these genes have associated pseudogenes [100,101]. In fact, NANOG has eleven pseudogenes [102].

The biological significance of these pseudogenes remains undetermined for the most part, but some are translated and expressed in somatic tissues, and absent translation, their transcripts can regulate the expression of the associated gene homologues. Ambady et al. [103] while attempting to generate a NANOG FP reporter cell line found that they had created two, one for "embryonic" NANOG (eNANOG) and one for a NANOG pseudogene (NANOGP8). They evaluated NANOG expression patterns in human stem cells and non-pluripotent cells and found evidence of NANOGP8 protein expression and translocation to nucleus where NANOGP8 protein likely functions as a transcription factor in non-pluripotent cells. They report that eNANOG, in contrast, is only expressed in pluripotent cells. Hawkins and Morris [104] studied the expression pattern of OCT4 and its pseudogenes and found evidence that OCT4 expression is regulated by non-coding pseudogene transcripts through its action as a "sink" for antisense-regulating siRNA.

The presence of pseudogenes is an important consideration when designing genome edits simply because current genome editing technologies depend on sequence homology for their targeting specificity. The presence of pseudogenes with high sequence homology therefore becomes a critical design consideration. The genome targeting design should specifically target the intended gene rather than any of the pseudohomologues; care must be taken in design of both the guide RNAs and the homology arms of the donor plasmid. In the case of NANOG, using the standard Cas9 enzyme, it is practically impossible to choose a guide RNA that does not potentially target many other members of the gene family. Thus, the ability to screen for the desired target location in the resulting cell lines becomes essential. Screening is complicated by the fact that genetic evaluation of copy number and confirmation of proper gene contextual insertion both depend on PCR which is itself driven by DNA sequence homology of the primers. Design of screening sequences is critical and should be part of the design of the editing strategy, and they should be reported so that specificity for the intended edit can be evaluated.

In addition, the confusion that result from the presence of pseudogenes, whether expressed as transcripts alone or as translated peptides, has been noted by several studies. Xu et al. [105] investigated expression patterns of OCT4 and its pseudogenes with an

emphasis on how the sequence similarity of these genes has caused confusion and misinterpretation of OCT4 expression. Their evaluation emphasized the importance of specificity of primer design and led to the identification of specific gene products (OCT4 or its pseudogenes) with detection by specific antibodies. Jez et al. [106] were able to differentiate OCT4 P1,2,4 but were forced to use elaborate methodologies such as RT-PCR followed by pseudogene-specific restriction digestion, cloning, and sequencing to distinguish between them and the "real" OCT4 gene product. These pseudogenes can be biologically important as there is evidence that expression of OCT4 pseudogenes, in addition to OCT4, follow a developmentally regulated pattern in differentiating human embryonic stem cells (hESCs). It will be critical for researchers to keep these caveats in mind when designing reporter cell lines and when drawing conclusions from their studies.

5. Conclusions

Advances in gene editing techniques, imaging methodologies, and modeling approaches are enabling the broader use of live cell imaging in studies of GRNs. The importance of the unique information that live cell imaging can add to omics data provide an incentive to overcome the associated challenges.

Disclaimer: Certain commercial equipment, instruments, or materials are identified in this chapter to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Briefings Bioinf 2016;19(2):286–302.
[2] Belluschi S, Calderbank EF, Ciaurro V, Pijuan-Sala B, Santoro A, Mende N, et al. Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. Nat Commun 2018;9(1):4100.
[3] Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 2012;150(6):1209–22.
[4] Prill RJ, Vogel R, Cecchi GA, Altan-Bonnet G, Stolovitzky G. Noise-driven causal inference in biomolecular networks. PLoS ONE 2015;10(6):e0125777.
[5] Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci U S A 2018;115(10):E2467–76.
[6] Tkačik G, Bialek W. Diffusion, dimensionality, and noise in transcriptional regulation. Phys Rev E 2009;79(5):051901.
[7] Filipczyk A, Marr C, Hastreiter S, Feigelman J, Schwarzfischer M, Hoppe PS, et al. Network plasticity of pluripotency transcription factors in embryonic stem cells. Nat Cell Biol 2015;17(10):1235–46.
[8] Dunn SJ, Martello G, Yordanov B, Emmott S, Smith AG. Defining an essential transcription factor program for naive pluripotency. Science 2014;344 (6188):1156–60.
[9] Rizzino A, Wuebben EL. Sox2/Oct4: A delicately balanced partnership in pluripotent stem cells and embryogenesis. Biochim Biophys Acta 2016;1859 (6):780–91.
[10] 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(47):19262–7.
[11] Cohen AA, Kalisky T, Mayo A, Geva-Zatorsky N, Danon T, Issaeva I, et al. Protein dynamics in individual human cells: experiment and theory. PLoS ONE 2009;4(4):e4901.
[12] Halter M, Sisan DR, Chalfoun J, Stottrup BL, Cardone A, Dima AA, et al. Cell cycle dependent TN-C promoter activity determined by live cell imaging. Cytometry A 2011;79A(3):192–202.
[13] Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, et al. Variability and memory of protein levels in human cells. Nature 2006;444(7119):643–6.
[14] Wolff SC, Kedziora KM, Dumitru R, Dungee CD, Zikry TM, Beltran AS, et al. Inheritance of OCT4 predetermines fate choice in human embryonic stem cells. Mol Syst Biol 2018;14(9):e8140.
[15] Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell 2014;55(2):319–31.
[16] Figge MT, Murphy RF. Image-based systems biology. Cytometry A 2015;87 (6):459–61.
[17] Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 2007;1(1):55–70.
[18] Costa MR, Ortega F, Brill MS, Beckervordersandforth R, Petrone C, Schroeder T, et al. Continuous live imaging of adult neural stem cell division and lineage progression in vitro. Development 2011;138(6):1057–68.
[19] Herberg M, Zerjatke T, de Back W, Glauche I, Roeder I. Image-based quantification and mathematical modeling of spatial heterogeneity in ESC colonies. Cytometry A 2015;87(6):481–90.
[20] Ravin R, Hoeppner DJ, Munno DM, Carmel L, Sullivan J, Levitt DL, et al. Potency and fate specification in CNS stem cell populations in vitro. Cell Stem Cell 2008;3(6):670–80.
[21] Strasser MK, Hoppe PS, Loeffler D, Kokkaliaris KD, Schroeder T, Theis FJ, et al. Lineage marker synchrony in hematopoietic genealogies refutes the PU.1/GATA1 toggle switch paradigm. Nat Commun 2018;9(1):2697.
[22] Kalmar T, Lim C, Hayward P, Munoz-Descalzo S, Nichols J, Garcia-Ojalvo J, et al. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol 2009;7(7):e1000149.
[23] Chen MJ, Lummertz da Rocha E, Cahan P, Kubaczka C, Hunter P, Sousa P, et al. Transcriptome dynamics of hematopoietic stem cell formation revealed using a combinatorial Runx1 and Ly6a reporter system. Stem Cell Rep 2020;14 (5):956–71.
[24] Den Hartogh SC, Schreurs C, Monshouwer-Kloots JJ, Davis RP, Elliott DA, Mummery CL, et al. Dual reporter MESP1 mCherry/w-NKX2-5 eGFP/w hESCs enable studying early human cardiac differentiation. Stem Cells 2015;33 (1):56–67.
[25] Faddah DA, Wang H, Cheng AW, Katz Y, Buganim Y, Jaenisch R. Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs. Cell Stem Cell 2013;13 (1):23–9.
[26] Wu J, Matthias N, Lo J, Ortiz-Vitali JL, Shieh AW, Wang SH, et al. A Myogenic double-reporter human pluripotent stem cell line allows prospective isolation of skeletal muscle progenitors. Cell Rep 2018;25(7):1966–1981 e1964.
[27] Zhang JZ, Termglinchan V, Shao NY, Itzhaki I, Liu C, Ma N, et al. A human iPSC double-reporter system enables purification of cardiac lineage subpopulations with distinct function and drug response profiles. Cell Stem Cell 2019;24(5):802–811 e805.
[28] Smith RCG, Stumpf PS, Ridden SJ, Sim A, Filippi S, Harrington HA, et al. Nanog fluctuations in embryonic stem cells highlight the problem of measurement in cell biology. Biophys J 2017;112(12):2641–52.
[29] Hubbard JB, Halter M, Sarkar S, Plant AL. The role of fluctuations in determining cellular network thermodynamics. PLoS ONE 2020;15(3): e0230076.
[30] Jost AP, Waters JC. Designing a rigorous microscopy experiment: validating methods and avoiding bias. J Cell Biol 2019;218(5):1452–66.
[31] Skylaki S, Hilsenbeck O, Schroeder T. Challenges in long-term imaging and quantification of single-cell dynamics. Nat Biotechnol 2016;34(11):1137–44.
[32] Wessels JT, Pliquett U, Wouters FS. Light-emitting diodes in modern microscopy–from David to Goliath?. Cytometry A 2012;81(3):188–97.
[33] Beier HT, Ibey BL. Experimental comparison of the high-speed imaging performance of an EM-CCD and sCMOS camera in a dynamic live-cell imaging test case. PLoS ONE 2014;9(1):e84614.
[34] Phillips ZF, Dean S, Recht B, Waller L. High-throughput fluorescence microscopy using multi-frame motion deblurring. Biomed Opt Express 2020;11(1):281–300.
[35] Chalfoun J, Kociolek M, Dima A, Halter M, Cardone A, Peskin A, et al. Segmenting time-lapse phase contrast images of adjacent NIH 3T3 cells. J Microsc 2013;249(1):41–52.
[36] Chalfoun J, Majurski M, Dima A, Halter M, Bhadriraju K, Brady M. Lineage mapper: a versatile cell and particle tracker. Sci Rep 2016;6:36984.
[37] Bhadriraju K, Halter M, Amelot J, Bajcsy P, Chalfoun J, Vandecreme A, et al. Large-scale time-lapse microscopy of Oct4 expression in human embryonic stem cell colonies. Stem Cell Res 2016;17(1):122–9.
[38] Hilsenbeck O, Schwarzfischer M, Skylaki S, Schauberger B, Hoppe PS, Loeffler D, et al. Software tools for single-cell tracking and quantification of cellular and molecular properties. Nat Biotechnol 2016;34(7):703–6.
[39] Bajcsy PC, Simon M. Web analysis of big image data. Springer International Publishing; 2018.
[40] Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, Martone ME, et al. Biological imaging software tools. Nat Methods 2012;9(7):697–710.
[41] Smith K, Piccinini F, Balassa T, Koos K, Danka T, Azizpour H, et al. Phenotypic image analysis software tools for exploring and understanding big image data from cell-based assays. Cell Syst 2018;6(6):636–53.
[42] Winter M, Mankowski W, Wait E, Temple S, Cohen AR. LEVER: software tools for segmentation, tracking and lineaging of proliferating cells. Bioinformatics 2016;32(22):3530–1.

[43] Bove A, Gradeci D, Fujita Y, Banerjee S, Charras G, Lowe AR. Local cellular neighborhood controls proliferation in cell competition. Mol Biol Cell 2017;28(23):3215–28.
[44] Maioli V, Chennell G, Sparks H, Lana T, Kumar S, Carling D, et al. Time-lapse 3-D measurements of a glucose biosensor in multicellular spheroids by light sheet fluorescence microscopy in commercial 96-well plates. Sci Rep 2016;6:37777.
[45] Voleti V, Patel KB, Li W, Perez Campos C, Bharadwaj S, Yu H, et al. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. Nat Methods 2019;16(10):1054–62.
[46] Liu Z, Legant WR, Chen BC, Li L, Grimm JB, Lavis LD, et al. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. Elife 2014;3:e04236.
[47] Netten H, van Vliet LJ, Boddeke FR, de Jong P, Young IT. A fast scanner for fluorescence microscopy using a 2-D CCD and time delayed integration. Bioimaging 1994;2(4):184–92.
[48] Schenk FW, Brill N, Marx U, Hardt D, Konig N, Schmitt R. High-speed microscopy of continuously moving cell culture vessels. Sci Rep 2016;6:34038.
[49] Zheng G, Horstmeyer R, Yang C. Wide-field, high-resolution Fourier ptychographic microscopy. Nat Photonics 2013;7(9):739–45.
[50] Tian L, Liu Z, Yeh L-H, Chen M, Zhong J, Waller L. Computational illumination for high-speed in vitro Fourier ptychographic microscopy. Optica 2015;2 (10):904–11.
[51] Kasprowicz R, Suman R, O'Toole P. Characterising live cell behaviour: traditional label-free and quantitative phase imaging approaches. Int J Biochem Cell Biol 2017;84:89–95.
[52] van der Graaff L, Boyaval F, van Vliet L, Stallinga S: Fluorescence imaging for whole slide scanning using LED-based color sequential illumination, vol. 10679: SPIE; 2018.
[53] Gupta A, Harrison PJ, Wieslander H, Pielawski N, Kartasalo K, Partel G, et al. Deep learning in image cytometry: a review. Cytometry A 2019;95 (4):366–80.
[54] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. arXiv 2015.
[55] Ling C, Majurski M, Halter M, Stinson J, Plant AL, Chalfoun J. Analyzing U-Net robustness for single cell nucleus segmentation from phase contrast images. The IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops, 2020.
[56] Ounkomol C, Seshamani S, Maleckar MM, Collman F, Johnson GR. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. Nat Methods 2018;15(11):917–20.
[57] Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, et al. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. PLoS Comput Biol 2016;12(11):e1005177.
[58] Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature 2006;440(7082):358–62.
[59] Banerji CR, Miranda-Saavedra D, Severini S, Widschwendter M, Enver T, Zhou JX, et al. Cellular network entropy as the energy potential in Waddington's differentiation landscape. Sci Rep 2013;3:3039.
[60] MacArthur BD, Lemischka IR. Statistical mechanics of pluripotency. Cell 2013;154(3):484–9.
[61] Antebi YE, Nandagopal N, Elowitz MB. An operational view of intercellular signaling pathways. Curr Opin Syst Biol 2017;1:16–24.
[62] Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J. Origins of extrinsic variability in eukaryotic gene expression. Nature 2006;439 (7078):861–4.
[63] Dunlop MJ, Cox 3rd RS, Levine JH, Murray RM, Elowitz MB. Regulatory activity revealed by dynamic correlations in gene expression noise. Nat Genet 2008;40(12):1493–8.
[64] Farkash-Amar S, Eden E, Cohen A, Geva-Zatorsky N, Cohen L, Milo R, et al. Dynamic proteomics of human protein level and localization across the cell cycle. PLoS ONE 2012;7(11):e48722.
[65] Chickarmane V, Olariu V, Peterson C. Probing the role of stochasticity in a model of the embryonic stem cell: heterogeneous gene expression and reprogramming efficiency. BMC Syst Biol 2012;6:98.
[66] Cox CD, McCollum JM, Allen MS, Dar RD, Simpson ML. Using noise to probe and characterize gene circuits. Proc Natl Acad Sci U S A 2008;105 (31):10809–14.
[67] Li C, Balazsi G. A landscape view on the interplay between EMT and cancer metastasis. NPJ Syst Biol Appl 2018;4:34.
[68] Pakka VH, Prugel-Bennett A, Dasmahapatra S. Correlated fluctuations carry signatures of gene regulatory network dynamics. J Theor Biol 2010;266 (3):343–57.
[69] Yu P, Nie Q, Tang C, Zhang L. Nanog induced intermediate state in regulating stem cell differentiation and reprogramming. BMC Syst Biol 2018;12(1):22.
[70] Gillespie DT. Exact stochastic simulation of coupled chemical-reactions. J Phys Chem-Us 1977;81(25):2340–61.
[71] Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology?. BioEssays 2012;34 (2):149–57.
[72] Tse MJ, Chu BK, Gallivan CP, Read EL. Rare-event sampling of epigenetic landscapes and phenotype transitions. PLoS Comput Biol 2018;14(8): e1006336.
[73] Waddington C. The strategy of the genes. In.: Unwin, London; 1957.
[74] Wang J, Xu L, Wang E. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. Proc Natl Acad Sci U S A 2008;105(34):12271–6.
[75] Zhang B, Wolynes PG. Stem cell differentiation as a many-body problem. Proc Natl Acad Sci U S A 2014;111(28):10185–90.
[76] Brackston RD, Wynn A, Stumpf MPH. Construction of quasipotentials for stochastic dynamical systems: an optimization approach. Phys Rev E 2018;98 (2–1):022136.
[77] Kang X, Li C. Landscape inferred from gene expression data governs pluripotency in embryonic stem cells. Comput Struct Biotechnol J 2020;18:366–74.
[78] Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 2008;453(7194):544–7.
[79] Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 2011;146(4):633–44.
[80] Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. Nature 2016;535(7611):299–302.
[81] Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells. Nat Biotechnol 2003;21(3):319–21.
[82] Gaj T, Gersbach CA, Barbas 3rd CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 2013;31(7):397–405.
[83] Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 2013;14(1):49–55.
[84] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339(6121):819–23.
[85] Hruscha A, Krawitz P, Rechenberg A, Heinrich V, Hecht J, Haass C, et al. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. Development 2013;140(24):4982–7.
[86] Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, et al. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature 2009;462(7273):595–601.
[87] Koch B, Nijmeijer B, Kueblbeck M, Cai Y, Walther N, Ellenberg J. Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing. Nat Protoc 2018;13(6):1465–87.
[88] Nakamura S, Maruyama A, Kondo Y, Kano A, De Sousa OM, Iwahashi M, et al. Asymmetricity between sister cells of pluripotent stem cells at the onset of differentiation. Stem Cells Dev 2018;27(5):347–54.
[89] Nur Patria Y, Lilianty J, Elefanty AG, Stanley EG, Labonne T, Bateman JF, et al. Generation of a SOX9-tdTomato reporter human iPSC line, MCRIi001-A-2, using CRISPR/Cas9 editing. Stem Cell Res 2020;42:101689.
[90] Sokolik C, Liu Y, Bauer D, McPherson J, Broeker M, Heimberg G, et al. Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. Cell Syst 2015;1(2):117–29.
[91] Teshigawara R, Hirano K, Nagata S, Ainscough J, Tada T. OCT4 activity during conversion of human intermediately reprogrammed stem cells to iPSCs through mesenchymal-epithelial transition. Development 2016;143 (1):15–23.
[92] Wray J, Kalkan T, Gomez-Lopez S, Eckardt D, Cook A, Kemler R, et al. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol 2011;13(7):838–45.
[93] Filipczyk A, Gkatzis K, Fu J, Hoppe PS, Lickert H, Anastassiadis K, et al. Biallelic expression of nanog protein in mouse embryonic stem cells. Cell Stem Cell 2013;13(1):12–3.
[94] Tiyaboonchai A, Mac H, Shamsedeen R, Mills JA, Kishore S, French DL, et al. Utilization of the AAVS1 safe harbor locus for hematopoietic specific transgene expression and gene knockdown in human ES cells. Stem Cell Res 2014;12(3):630–7.
[95] Hatano SY, Tada M, Kimura H, Yamaguchi S, Kono T, Nakano T, et al. Pluripotential competence of cells associated with Nanog activity. Mech Dev 2005;122(1):67–79.
[96] Roberts B, Hendershott MC, Arakaki J, Gerbin KA, Malik H, Nelson A, et al. Fluorescent gene tagging of transcriptionally silent genes in hiPSCs. Stem Cell Rep 2019;12(5):1145–58.
[97] Di Giammartino DC, Apostolou E. The chromatin signature of pluripotency: establishment and maintenance. Curr Stem Cell Rep 2016;2(3):255–62.
[98] Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, et al. 3D chromosome regulatory landscape of human pluripotent cells. Cell Stem Cell 2016;18(2):262–75.
[99] Chatterjee S, Stochaj U. Diffusion of proteins across the nuclear envelope of HeLa cells. Biotechniques 1998;24(4):668–74.
[100] Clark AT, Rodriguez RT, Bodnar MS, Abeyta MJ, Cedars MI, Turek PJ, et al. Human STELLAR, NANOG, and GDF3 genes are expressed in pluripotent cells and map to chromosome 12p13, a hotspot for teratocarcinoma. Stem Cells 2004;22(2):169–79.
[101] Pain D, Chirn GW, Strassel C, Kemp DM. Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. J Biol Chem 2005;280(8):6265–8.
[102] Booth HA, Holland PW. Eleven daughters of NANOG. Genomics 2004;84 (2):229–38.

[103] Ambady S, Malcuit C, Kashpur O, Kole D, Holmes WF, Hedblom E, et al. Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells. Int J Dev Biol 2010;54(11–12):1743–54.
[104] Hawkins PG, Morris KV. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Transcription 2010;1(3):165–75.
[105] Xu G, Yang L, Zhang W, Wei X. All the tested human somatic cells express both Oct4A and its pseudogenes but express Oct4A at much lower levels compared with its pseudogenes and human embryonic stem cells. Stem Cells Dev 2015;24(13):1546–57.
[106] Jez M, Ambady S, Kashpur O, Grella A, Malcuit C, Vilner L, et al. Expression and differentiation between OCT4A and its Pseudogenes in human ESCs and differentiated adult somatic cells. PLoS ONE 2014;9(2):e89546.