Published in IET Systems Biology
Received on 2nd February 2009
Revised on 10th April 2009
doi: 10.1049/iet-syb.2009.0011

ISSN 1751-8849

Systems biology approaches to understanding stem cell fate choice

J. Peltier[1] *D.V. Schaffer*[2]

[1]*Department of Chemical Engineering, University of California, 176 Stanley Hall, Berkeley, CA, 94720-3220, USA*
[2]*Department of Chemical Engineering, Department of Bioengineering, and The Helen Wills Neuroscience Institute, University of California, 176 Stanley Hall, Berkeley, CA, 94720-3220, USA*
E-mail: schaffer@berkeley.edu

Abstract: Stem cells have the capability to self-renew and maintain their undifferentiated state or to differentiate into one or more specialised cell types. Stem cell expansion and manipulation *ex vivo* is a promising approach for engineering cell replacement therapies, and endogenous stem cells represent potential drugable targets for tissue repair. Before we can harness stem cells' therapeutic potential, we must first understand the intracellular mechanisms controlling their fate choices. These mechanisms involve complex signal transduction and gene regulation networks that feature, for example, intricate feed-forward loops, feedback loops and cross-talk between multiple signalling pathways. Systems biology applies computational and experimental approaches to investigate the emergent behaviour of collections of molecules and strives to explain how these numerous components interact to regulate molecular, cellular and organismal behaviour. Here we review systems biology, and in particular computational, efforts to understand the intracellular mechanisms of stem cell fate choice. We first discuss deterministic and stochastic models that synthesise molecular knowledge into mathematical formalism, enable simulation of important system behaviours and stimulate further experimentation. In addition, statistical analyses such as Bayesian networks and principal components analysis (PCA)/partial least squares (PLS) regression can distill large datasets into more readily managed networks and principal components that provide insights into the critical aspects and components of regulatory networks. Collectively, integrating modelling with experimentation has strong potential for enabling a deeper understanding of stem cell fate choice and thereby aiding the development of therapies to harness stem cells' therapeutic potential.

1 Introduction

Stem cells – first discovered in mouse bone marrow by Becker *et al.* [1] and Till and McCulloch [2] – are defined by their two hallmark properties: (i) self-renewal, or extended maintenance and potentially proliferation in an undifferentiated state, and (ii) differentiation into one or more specialised cell types. Pluripotent embryonic stem (ES) cells can give rise to any cell type in an adult organism, whereas multipotent adult stem cells are capable of generating a more limited set of cell types, typically ones in the tissue in which stem cells reside. Their ability to self-renew or differentiate into multiple cell types makes stem cells promising therapeutic candidates in cell replacement therapies for multiple injuries and diseases, including diabetes [3], spinal cord injury [4] and Alzheimer's disease [5], among others. However, before we can harness stem cells' therapeutic potential and guide their production of a desired cell type, we must first identify the factors and understand the mechanisms that govern their behaviour and fate choices.

Whether in a multi-cellular organism or a culture dish, a stem cell constantly receives environmental cues in many forms: soluble cues from proteins such as mitogens and cytokines [6–9], small molecules [10, 11], and nutrients [12, 13], as well as 'solid phase' cues such as cell–cell contacts and the biochemical and mechanical properties of

the extracellular matrix, including signals immobilised to it [14–17]. These signals guide the stem cell towards specific behaviours, such as survival, apoptosis, self-renewal or differentiation into one of multiple lineages (Fig. 1). An instructive view of stem cell fate choice states that environmental cues initiate the intracellular signals that direct the cells to their fate, whereas a selective mechanism indicates that environmental factors merely support the survival of certain fates. It appears likely that both of these modes operate in different tissues and circumstances [18]. Regardless of which mechanism is operating, however, stem cell behaviour is guided by molecular interactions and reactions involving receptors, signalling networks and transcription factors. In particular, signal processing networks that relay input signals from the cell surface to the nucleus feature complex, non-linear components such as feed-forward and feedback loops, signal amplification cascades and cross-talk between multiple signalling pathways. Information processing continues within the nucleus where transcription factor networks control the expression of themselves and each other, in addition to the target genes required for execution of fate choice. The result is a complex, multi-level, non-linear system that can exhibit a number of rich behaviours, including switches and oscillations [19–22]. These behaviours are critical regulators of stem cell self-renewal and differentiation, and in many ways they are difficult to investigate and interpret intuitively without the aid of systems-level analysis and the accompanying mathematical tools. A topic closely related to stem cell fate choice, organismal development, is also studied by systems biologists. However, we will focus on the molecular mechanisms of fate choice within single cells and refer readers to an excellent review of models of multi-cellular pattern formation [23].

Systems biology is a field that studies the collective behaviour of groups of complex, interacting biological components. Its approaches offer advantages that complement and enhance traditional reductionist experimental avenues that tend to focus more on individual components than on interactions occurring within a larger scale system. Systems biology analyses of large biological systems such as cells often rely on computational models, which serve many uses: (i) they summarise our knowledge of and assumptions about a system into formal, mathematical statements; (ii) they highlight gaps in our knowledge of a system; (iii) they generate hypotheses about the behaviour of the system that motivate experimentation

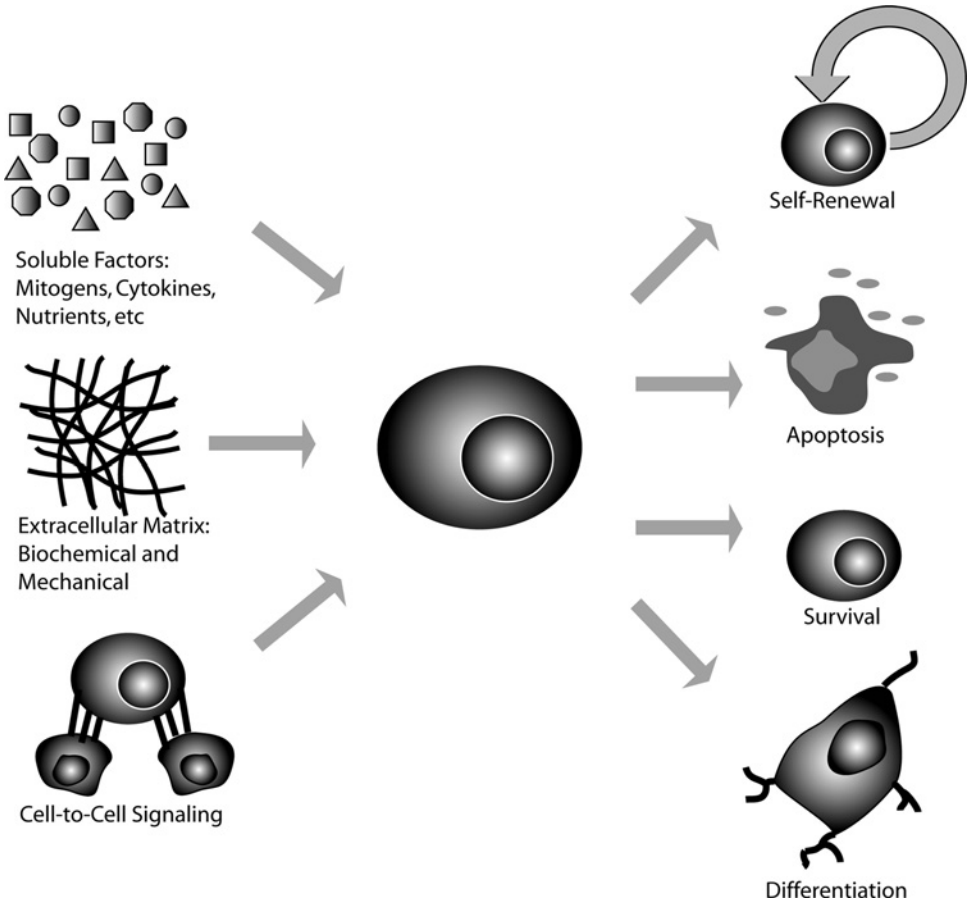

Figure 1 *A stem cell as a signal processor*

A stem cell receives multiple signals from its environment: soluble factors such as mitogens, cytokines, and nutrients; the biochemical and mechanical properties of the extra cellular matrix; and contacts from neighbouring cells. The cell processes these signals to determine its fate: self-renewal, survival, apoptosis, or differentiation

doi: 10.1049/iet-syb.2009.0011

and further modelling; (iv) they aid in the analysis of large datasets – such as those generated by genomic, transcriptomic, proteomic and kinomic experimentation – and thus summarise the data and highlight important, potentially unintuitive behaviour for future experimentation [24–27], and (v) they highlight critical loci within a system that can be manipulated to generate a desired outcome. As examples of the last point, a model can be used to pinpoint drugable therapeutic targets that direct endogenous stem cell pools to a desired fate or to enable the design of strategies to optimise *ex vivo* expansion for cellular therapy [28, 29].

In this review, we discuss systems biology modelling approaches and techniques that are increasingly utilised to understand the intracellular mechanisms of stem cell fate choice. Deterministic and stochastic computational models have formally synthesised our molecular knowledge into mathematical statements, furthered our understanding of important network behaviours and motivated future experimentation. Statistical analyses such as Bayesian networks and principal components analysis (PCA) have distilled large datasets into tractable candidate networks and principal components that are then used to derive insight into the critical pieces of the fate choice network. Collectively, these efforts have furthered the stem cell field and brought us closer to the eventual goal of harnessing the therapeutic promise of stem cells.

2 Deterministic models

A deterministic model, one that always yields the same result given the same set of initial conditions, often consists of a set of either ordinary differential equations (ODEs) or partial differential equations (PDEs) that typically model the mass action of molecular species. ODEs are used in situations where the system (i.e. the intracellular or extracellular region) is spatially homogeneous. This assumption can be relaxed somewhat by compartmentalising the system into several homogeneous sub-systems with transport between them. PDEs are used when spatial heterogeneity in species concentrations or other dependent variables becomes more complex and must be captured in order to accurately model the system. These ODE- or PDE-based models can be analysed at steady state or dynamically, yielding either algebraic equations or spatially dependent PDEs/ODEs, respectively.

Deterministic models are often used to simulate the intracellular chemical reactions and interactions regulating stem cell fate choice, including ligand–receptor dynamics, signal transduction pathways and transcription factor networks. To do so, the modeller must have a precise understanding of these constituent molecular interactions within the network and knowledge of their rate and binding constants. The latter in particular is often limiting, as in many systems there is a dearth of experimentally measured constants, often requiring estimation of constants based on analogous systems.

Models are most often solved numerically, and the resulting output is kinetic information about the quantities of network species as a function of time and/or at steady state. This enables an investigation of how different parameters that describe molecular interactions (e.g. binding or catalysis) affect system behaviour, driving the formation of new hypotheses. Additionally, models may be used to identify sensitive loci within the network, that is, locations where small perturbations can exert a strong effect on cell behaviour. These could represent 'drugable targets' for therapeutic intervention or failure points where natural mutations may adversely affect system function and lead to disease.

In an early effort to quantitatively understand the molecular basis for stem cell fate choice, a ligand–receptor signalling threshold (LIST) model was proposed [30, 31]. This model posited that a threshold level of ligand–receptor signalling is an important determinant of stem cell fate and was useful in predicting cellular responses to various cytokine concentrations. It even demonstrated that a cytokine's ability to maintain pluripotent ES cells was dependent on its receptor binding properties, such as heterodimerisation instead of homodimerisation. The intracellular signal transduction and gene regulation processes, however, were not treated within the scope of this model.

Several papers have used deterministic models to mathematically investigate the role of intracellular signal transduction and transcription factor networks in stem cell fate choice. One behaviour that emerges is network bistability, where as an input parameter is continuously varied, the system output transitions between one, then two, then one stable steady state solution (Fig. 2). The two bifurcation points, where the number of stable solutions transitions between one and two, represent quantitative input threshold levels where the system qualitatively switches state. Bifurcation thus serves as an analogue-to-digital converter to translate a graded input signal into an unambiguous, 'all or nothing' behavioural response (e.g. self-renewal instead of differentiation). The behaviour also serves a second important function. As discussed below (stochastic models), noise is a feature of many biological systems and bistable systems exhibit hysteresis, where for example, the input level at which the system switches from the first to the second state is higher than that at which it flips back from the second to the first (Fig. 2). Hysteresis thus filters the noise within a system to prevent potentially deleterious rapid switching between states that would otherwise result from noise in the input parameter or signal. The initial example of such behaviour in a stem cell regulatory network was the Sonic hedgehog (Shh) signalling system [19]. Shh, which patterns tissues during development and is an important mitogen for adult neural stem cells [6], drives expression of the transcription factor Gli1, which positively regulates its own transcription. Furthermore, Gli1 upregulates the expression of Patched, a repressor of Shh signalling. This nested positive and

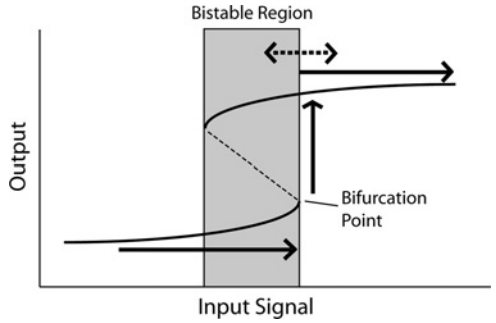

Figure 2 *Bistability acts as a cell fate switch*

In a bistable system, a range of input values generates two stable output solutions (the bistable region). When the input value increases to the bifurcation point the system switches state and 'jumps' to the upper portion of the curve. That is, the system converts the analogue input signal to a digital output (i.e. fate choice). Once the system has switched states, hysteresis allows it to robustly resist the effects of input noise (represented by dashed double arrow). The system is unable to 'jump' back to the lower curve unless the input signal decreases below the bistable region

negative feedback renders an intuitive understanding of the pathway behaviour difficult, and a Shh network model found that Gli1 expression exhibits bistability as a function of the input Shh concentration [19]. This outcome demonstrates one possible mechanism by which stem cells commit to a specific fate, that is, once Shh concentration exceeds a threshold, and Gli1 expression switches to a 'high' state, a moderate decrease in Shh will not return Gli1 expression to the 'low' state.

Several additional efforts have modelled transcription factor networks that exhibit bistability [20, 21]. Chickarmane *et al.* [20] modelled the interplay between three canonical pluripotency transcription factors crucial for ES cell self-renewal: Oct4, Sox2 and Nanog. Each of these genes positively regulates the expression of the others in addition to itself, as well as downstream target genes that either maintain pluripotency or induce differentiation [32–34]. This network gives rise to bistable expression of Oct4, Sox2 and Nanog, leading to a plausible mechanistic explanation for pluripotency maintenance. Since the publication of this work, however, other work has discovered additional transcription factors that play a role in maintaining pluripotency [35]. Future investigations of this updated network should prove interesting.

One genetic switch in hematopoietic stem cells involves the transcription factors GATA-1 and PU.1. Low GATA-1 and PU.1 expression maintain the cell in an undifferentiated state, whereas dominant expression of GATA-1 promotes the erythroid/magakaryocyte lineage, and PU.1 promotes the myeloid lineage. GATA-1 and PU.1 both stimulate their own transcription and inhibit that of the other, resulting in a network that generates a bistable, genetic toggle switch [21]. Huang *et al.* [36] examined the same GATA-1/PU.1 system to determine how a cell transitions from a progenitor state to one of the two possible differentiated progeny. They used a simple model of the GATA-1/PU.1 network, consisting of two ODEs, to generate what they termed an 'attractor' landscape, which is loosely analogous to a potential landscape where the stable cell states (progenitors or their differentiated progeny) reside in alternate 'wells'. Using this model, they could investigate how the shape of this landscape varies with levels of transcription factor auto-stimulation, cross-inhibition and decay. They then analysed possible mechanisms by which the landscape could be altered by the changes in transcription factor expression and activity that are triggered by differentiation signals. They found that the paradigm that resulted in the best match between the model results and experimental measurements of mRNA levels was one in which a differentiation signal altered the transcription factor landscape such that the 'progenitor point' was no longer in a valley but rather on a peak. This rendered the progenitor state unstable and forced the cell into a stable differentiated state. The authors found that this landscape shift could be accomplished by a reduction in transcription factor auto-stimulation, an increase in transcription factor decay, or both.

In addition to bistability, recent work has shown that network oscillations may also play a role in stem cell fate choice, specifically in the maintenance of adult neural stem cells by Notch signalling [22] Activation of the Notch pathway stimulates expression of the Hes family of transcription factors, which then inhibit their own transcription. This pathway can act as a switch that is important for developmental pattern formation [37, 38] or as an oscillator [39, 40] that is important for stem cell maintenance [22]. We have recently developed a single mathematical model that demonstrates that the Notch network can operate as a switch or an oscillator depending on the value of one key parameter. Specifically, tuning a single factor – the extent to which Hes1 binding reduces its expression – causes the network to transition between functioning as a bistable switch and an oscillator [41].

There have been additional efforts to mathematically model common signalling pathways downstream of key growth factors and cytokines [42, 43]. These include the effects of neurotrophin-3 (NT-3) on the MAPK pathway in ES cell-derived neural progenitors. The results provided insight into threshold levels of NT-3 stimulation and MAPK activity required for neuronal differentiation [42]. A similar study quantitatively studied the intracellular response to the cytokine leukemia inhibitory factor (LIF), which is crucial for murine ES cell self-renewal. This work uncovered a likely positive feedback loop that stimulated production of components of the LIF signalling pathway during LIF signalling; however, removal of LIF causes lower expression of LIF signalling factors, thereby decreasing the cell's overall sensitivity to LIF. The authors demonstrated that this desensitisation was a precursor to differentiation and loss of ES cell markers [43].

doi: 10.1049/iet-syb.2009.0011

Furthermore, a systems-level analysis of stem cell fate choice can highlight potentially non-intuitive therapeutic targets for stem cell control and enable optimisation of ex vivo stem cell production (as reviewed [28]). Zandstra and colleagues [29] developed a mass action model of the JAK/STAT3 pathway, which stimulates murine ES cell self-renewal. Upon binding of LIF to the LIF receptor (LIFR) and glycoprotein-130 (gp130), the resulting receptor complex triggers signalling and leads to phosphorylation and activation of the transcription factor STAT3. The authors experimentally verified the prediction that gp130 overexpression actually decreases STAT3 activation, because the excess gp130 binds and sequesters LIFR into non-signalling heterodimers. Additionally, a sensitivity analysis predicted and experimental results confirmed the additive importance of two individually 'innocuous' parameters (STAT3 nuclear export rate and JAK-mediated receptor activation rate), demonstrating the utility of such a model in generating novel hypotheses and potential therapeutic interventions. Finally, as predicted by the model, continual LIF stimulation causes desensitisation to ligand stimulation. The model enabled the design and experimental validation of a LIF addition protocol to maximise STAT3 phosphorylation, representing a useful application of a model for maximising *ex vivo* production of pluripotent ES cells for potential therapeutic use.

Deterministic models of the intracellular mechanisms of stem cell fate choice aid in interpreting complex network interactions, highlight new research avenues and potential therapeutic interventions and can potentially be applied to improve process development efforts for *ex vivo* cell production. However, detailed knowledge of the signal transduction and transcription factor networks of interest is required, including quantitative knowledge of the kinetic and equilibrium constants involved. A further limitation of these models is their failure to account for the 'noisy' behaviour that can arise in biological systems because of slow chemical reactions and/or small numbers of molecules, and the next section focuses on efforts to take these effects into account when modelling stem cell fate choice.

3 Stochastic models

Ever since stem cells were first discovered in the hematopoietic system, researchers have been studying the role of stochastics in stem cell fate choice [44–46]. When investigating the intracellular mechanisms governing stem cell fate, stochastics become important when the regulatory networks involve slow biochemical reactions and/or a small number of constituent molecules (e.g. transcripts, proteins, second messengers etc.) within a cell. Since landmark theoretical work by McAdams and Arkin [47] sparked interest in stochastic effects in gene expression, a considerable amount of work has focused on the implications of noise in multiple biological processes (see review [48]), including phage infection [49], *Bacillus subtilis* stress response mechanisms [50], circadian rhythm control [51], lymphocyte activation [52] and others [53]. Stochastic effects are sometimes able to explain phenomena not predicted by deterministic models. In fact, some work indicates that precise circadian rhythms are actually dependent on noise [51]. Furthermore, a stochastic model based on T cell antigen response has described situations where a bimodal output is observed, whereas a deterministic model predicts only a single, intermediate solution [54].

To date, very little work has investigated stochasticity in intracellular stem cell signalling pathways. As mentioned above, our model of the Shh signalling system predicts that expression of the transcription factor Gli1 exhibits deterministic bistability as a function of an input Shh signal. We also implemented stochastic simulations to show that stochastic effects near bifurcation points can lead to random switching between states, undermining deterministic switch-like behaviour in the network (similar to subsequent observations in other systems [55]). However, in the Shh network, the effects of noise are moderated by the Gli1-driven expression of the Shh repressor Patched. Specifically, whereas positive feedback amplifies noise, this negative feedback loop functions to dampen noise, resulting a robust switch that reliably directs stem cell fate despite inherent stochastic effects [19].

There is also experimental evidence that stochastics may be important in networks that control stem cell behaviour. Recent studies show that populations of stem cells can exist in multiple metastable states, and that cells within the population are capable of switching between these states (reviewed in [56]). For instance, in one study approximately 80% of murine ES cultures expressed the transcription factor Nanog. This observation could be readily explained if the other 20% were differentiated; however, this same 80/20 distribution is re-established when high or low marker populations are separated and cultured in isolation. Furthermore, ES cells with high Nanog expression are less likely to differentiate than low Nanog expressers, indicating that the system is more complex than simple bimodal Nanog expression [57]. Similar switches are observed in hematopoietic stem cells where the surface marker Sca-1 is expressed in a broad distribution. Analogous to Nanog expression in ES cells, after being separated by flow cytometry, high or low Sca-1 expressing populations re-establish the original distribution within several population doublings. Additionally, low Sca-1 expressing cells preferentially differentiate into the erythroid lineage, whereas high Sca-1 expressers favor the myeloid lineage [58]. The mechanistic basis of these switches is not yet understood, but a mathematical model involving stochastic state transitions between multiple stable states indicates that stochastic gene expression likely plays a role [58]. Elucidating the true nature of these noisy transitions and their effects on fate choice will require further studies that utilise a systems-level approach to complement experimental investigations.

4 Bayesian networks

The above approaches to understanding stem cell fate choice are applicable when the signalling network of interest is relatively well understood (such as Shh or JAK/STAT3 signalling). However, in many situations, the underlying molecular interactions are not yet well characterised or in large part unknown, but systems biology approaches can potentially help elucidate such unknown networks through the analysis of large datasets. One such technique is Bayesian network analysis (for mathematical details, see recent reviews [59, 60]), which can help reverse engineer signal transduction cascades from an '−omic' dataset (proteomic, transcriptomic, kinomic etc.) and deduce candidate causal relationships between measured variables or quantities. The result is a graphical map of probable interactions, which provides a physical/biochemical interpretation of the dataset and aids in the formulation of hypotheses for future experimentation. The network analysis is probabilistic in that it treats each measurement (i.e. mRNA concentration, protein phophorylation level etc.) as an uncertain estimate and therefore incorporates measurement noise in a systematic way. It is important to note that connections between species represent a causal but not necessarily a direct biochemical relationship, that is, two connected species may not have a direct physical interaction but may instead be separated by several intermediate steps. Whether the analysis detects these intermediate steps is dependent on the quality, size and detail of the dataset being analysed. Another limitation of a Bayesian network is that it does not provide information on the stepwise progression of the interactions. Rather, it predicts the likelihood of finding a species in a particular state given the states of the surrounding species. Finally, although a traditional requirement is that Bayesian network structures are acyclic and are therefore unable to capture feedback, recent work has developed a new technique to circumvent this requirement and recover feedback loops within a signal cascade [61, 62].

Bayesian networks have been used to predict microRNA targets [63], regulatory relationships between genes [64], and the effects of single nucleotide polymorphisms on the clinical outcome of sickle cell anemia [65]; however, comparatively little work has utilised Bayesian networks for the analysis of stem cell fate control. A Bayesian network model has been used to analyse a large proteomic dataset from mouse ES cells containing measurements of the phosphorylation states of several signalling molecules under multiple cytokine and extracellular matrix conditions [24]. The resulting model provided good agreement with several previously described stem cell signalling pathways despite the complete lack of *a priori* assumptions regarding these signalling systems, including the LIF/JAK/STAT3 pathway and the MAPK/ERK pathway. Additionally, the model predicted several novel links. For instance, the rate of conversion from undifferentiated to differentiated cells was found to be most dependent on the phosphorylation states of Adducin α and ERK2. Furthermore, the model predicted that Raf1 and PKCε would exert an effect on differentiated cell growth. The authors went on to experimentally verify these results, demonstrating that Bayesian networks can provide useful insights into the complex biological processes underlying stem cell fate choice. However, the authors caution that care must be taken when implementing these models, because, like many other analyses, the output results depend strongly on the quality and breadth of input data [24]. For instance, the analysis predicted that the rate of undifferentiated cell proliferation depended on the cytokine LIF without depicting any intermediate steps, when in reality there are numerous molecular intermediates that were not measured in the original dataset.

As large datasets needed for meaningful model predictions – which are resource- and time-intensive to generate – become increasingly available, Bayesian network analysis may become more utilised in the stem cell field to make novel predictions and drive new hypotheses regarding the signalling events that control stem cell fate.

5 Principal components analysis and partial least squares regression

Other techniques to analyse large datasets include PCA and partial least squares (PLS) regression (reviewed in [66]). Each measured quantity within a dataset (e.g. phophoprotein and/or transcription factor concentrations) can be depicted as an axis within 'signalling space' – analogous to how time and concentration are typically the x- and y-axes when graphing data. The resulting data space can have dozens (perhaps hundreds) of axes or dimensions, one for each measured quantity. PCA reduces this large number of dimensions to a few new axes called principal components. Each principal component represents a combination of the original signalling axes that have high covariance with one other. This reduces the data space to just a few tractable dimensions, allowing the researcher to more easily search for trends within the data. PLS is an extension of PCA that generates a predictive relationship between independent and dependent principal components.

PCA has found uses in many biological applications, including analysing neuronal decision-making processes [67], microarray data [68], libraries of chemical inhibitors [69] and signal transduction pathways [70], in addition to stem cell fate choice [25, 26]. As one example, Sharov *et al.* [25] amassed a large collection (nearly 250 000) of expressed sequence tags (ESTs) from public and other sources to produce a database of the mouse transcriptome. They then analysed this database for differences in EST frequency between cell types of varying potency. PCA identified a principal component that effectively represented a cell's developmental potential, from totipotent oocytes to fully differentiated newborn tissues. The expression levels

doi: 10.1049/iet-syb.2009.0011

of a set of 88 genes were closely associated with this principal component axis. All 88 genes followed a general trend of decreased expression with increased differentiation. Similar work analysed global gene expression changes during neuronal differentiation [26]. This analysis – which included ES cells, adult neural stem cells and neurons – identified a principal component axis composed of several genes that described the cell's level of neuronal commitment. It seems likely that similar axes will be found for other tissues, and the genes associated with these principal components are potential targets of research into the molecular mechanisms of differentiation.

PLS, a predictive extension of PCA, only recently found use in biological research when it was first used as a predictor of apoptosis resulting from various molecular perturbations [71, 72]. It has since been used to generate a protein signature consistent with metastatic breast cancer [73] and to investigate the migration and proliferation of mammary epithelial cells [74]. Recently, PLS regression has been used to demonstrate that multiple cell types process upstream kinase signals through a similar 'effector-processing' system to generate cell-specific responses to the same extracellular stimulus [75]. PLS was also used to analyse murine ES cell fate choice [27]. Using results from the same dataset as Woolf *et al.* [24] (see Bayesian networks above), which included phospho-protein measurements from multiple signalling pathways under multiple culture conditions, they correlated phosphorylation state to ES cell differentiation and self-renewal. The analysis revealed a set of seven signalling molecules closely associated with ES cell fate choice, including PKCε. Additionally, the PLS-generated model predicted that PKCε inhibition would slow the proliferation of differentiated cells. To test this prediction, the authors experimentally inhibited PKCε activity and found that it did indeed inhibit proliferation in differentiated cells with little effect on undifferentiated cells. This effect of PKCε inhibition was potentially moderate, however, because it involved blocking only one component of a much larger network. At any rate, this work successfully utilised systems biology to highlight key contributors to ES cell differentiation and self-renewal, and similar analyses may in the future determine whether the same signalling molecules play the same roles in other stem cell types and aid the development of *ex vivo* expansion technologies and therapeutic strategies.

6 Summary and future directions

The molecular mechanisms of stem cell fate choice involve multiple, interacting signalling pathways and transcription factor networks. The resulting signal processing circuitry is extremely complex and difficult to investigate solely through reductionist approaches; therefore increasing numbers of efforts have pursued a systems approach to understanding stem cell fate choice. The resulting models provide considerable insight into stem cell fate choice (summarised in Table 1). Several models have highlighted the importance of bistability and switches [19–21], as well as oscillations [22, 41]. Modelling also highlights gaps in our understanding, generates new hypotheses about network function and behaviour, and highlights critical control points that can be manipulated for cellular

Table 1 Summary of the model types used for analysing stem cell fate choice, and the key studies that employed these models

	Deterministic	Stochastic	Bayesian	PCA/PLS
Advantages	• insight into complex networks • aids therapeutic/process development	• systems with low numbers of molecules and/or slow biochemical reactions • behaviours not predicted by deterministic models	• reverse engineers network structure with no *a priori* knowledge	• condenses large dataset to several key parameters • predicts network outputs with no knowledge *a priori* of network structure
Requirements	• detailed knowledge of pathway, including kinetic data	• detailed knowledge of pathway	• large, high-quality datasets	• large, high-quality datasets
Key stem cell studies	• bistable switches in Shh and transcription factor networks [19–21] • switches and oscillations in Notch signalling [41] • common signalling pathways [42, 43] • optimisation of *ex vivo* cell production [29]	• stochastic effects near bifurcation points [19] • stochastic switching between multiple metastable states [58]	• analysis of proteomic data from mouse ES cells to highlight novel network links [24]	• analysis of transcriptome data from cells of varying potency [25, 26] • murine ES cell fate choice from analysis of phospho-protein data [27]

expansion and control. Network manipulation for cellular control has already met with success in some stem cell types [28, 29], and we anticipate these successes will continue as regulatory mechanisms of other stem cell types is subjected to systems analysis. Another interesting avenue of future research is how stochastics affect stem cell function, both for endogenous stem cell behaviour during development and adulthood, as well as in culture when extraction of a stem cell from its 'comfortable' niche may render it more susceptible to stochastic behaviour. There is evidence that the expression levels of some key genes that control stem cell self-renewal fluctuate considerably, and that the underlying mechanisms governing these switches are susceptible to stochastic effects [56–58]. Complementary experimental and modelling studies may yield further insights into this apparent randomness in the regulation of stem cell behaviour.

Advances in high-throughput experimental techniques have created large −omic datasets that are difficult to interpret without statistical analyses such as Bayesian networks and/or PCA/PLS. Some work has already utilised these techniques to highlight novel molecular interactions and key genes regulating stem cell fate [24–27], and future application should help further our understanding of these systems. The role of these candidates can then be tested experimentally to expand our knowledge of fate choice mechanisms. Furthermore, statistical analyses performed to date have primarily focused on murine ES cells, and future analyses in other stem cell types such as human pluripotent stem cells should prove equally fruitful.

In closing, the application of systems biology to the problem of stem cell fate choice is still young, and opportunities abound. Collectively, these efforts will bring us closer to a molecular understanding of stem cell fate choice and may aid the development of therapies for many debilitating injuries and diseases.

7 Acknowledgments

This work was supported by NIH R21EB007295, NIH R21DE018044 and a training grant fellowship (to JP) from the California Institute for Regenerative Medicine (Training Grant Number T1-00007).

8 References

[1] BECKER A.J., MCCULLOCH E.A., TILL J.E.: 'Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells', *Nature*, 1963, **197**, pp. 452–454

[2] TILL J.E., MCCULLOCH E.A.: 'A direct measurement of the radiation sensitivity of normal mouse bone marrow cells', *Radiat. Res.*, 1961, **14**, pp. 213–222

[3] ZHOU Q., BROWN J., KANAREK A., RAJAGOPAL J., MELTON D.A.: 'In vivo reprogramming of adult pancreatic exocrine cells to beta-cells', *Nature*, 2008, **455**, (7213), pp. 627–632

[4] BARNABE-HEIDER F., FRISEN J.: 'Stem cells for spinal cord repair', *Cell Stem Cell*, 2008, **3**, (1), pp. 16–24

[5] ZHAO C., DENG W., GAGE F.H.: 'Mechanisms and functional implications of adult neurogenesis', *Cell*, 2008, **132**, (4), pp. 645–660

[6] LAI K., KASPAR B.K., GAGE F.H., SCHAFFER D.V.: 'Sonic hedgehog regulates adult neural progenitor proliferation in vitro and in vivo', *Nat. Neurosci.*, 2003, **6**, (1), pp. 21–27

[7] SMITH A.G., HEATH J.K., DONALDSON D.D., *ET AL.*: 'Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides', *Nature*, 1988, **336**, (6200), pp. 688–690

[8] WILLIAMS R.L., HILTON D.J., PEASE S., *ET AL.*: 'Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells', *Nature*, 1988, **336**, (6200), pp. 684–687

[9] PALMER T.D., MARKAKIS E.A., WILLHOITE A.R., SAFAR F., GAGE F.H.: 'Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS', *J. Neurosci.*, 1999, **19**, (19), pp. 8487–8497

[10] TAKAHASHI J., PALMER T.D., GAGE F.H.: 'Retinoic acid and neurotrophins collaborate to regulate neurogenesis in adult-derived neural stem cell cultures', *J. Neurobiol.*, 1999, **38**, (1), pp. 65–81

[11] MADEN M.: 'Retinoic acid in the development, regeneration and maintenance of the nervous system', *Nat. Rev. Neurosci.*, 2007, **8**, (10), pp. 755–765

[12] DRUMMOND-BARBOSA D., SPRADLING A.C.: 'Stem cells and their progeny respond to nutritional changes during *Drosophila* oogenesis', *Dev. Biol.*, 2001, **231**, (1), pp. 265–278

[13] MORRISON S.J., CSETE M., GROVES A.K., MELEGA W., WOLD B., ANDERSON D.J.: 'Culture in reduced levels of oxygen promotes clonogenic sympathoadrenal differentiation by isolated neural crest stem cells', *J. Neurosci.*, 2000, **20**, (19), pp. 7370–7376

[14] METALLO C.M., MOHR J.C., DETZEL C.J., DE PABLO J.J., VAN WIE B.J., PALECEK S.P.: 'Engineering the stem cell microenvironment', *Biotechnol. Prog.*, 2007, **23**, (1), pp. 18–23

[15] ENGLER A.J., SEN S., SWEENEY H.L., DISCHER D.E.: 'Matrix elasticity directs stem cell lineage specification', *Cell*, 2006, **126**, (4), pp. 677–689

doi: 10.1049/iet-syb.2009.0011

[16] XU C., INOKUMA M.S., DENHAM J., *ET AL.*: 'Feeder-free growth of undifferentiated human embryonic stem cells', *Nat. Biotechnol.*, 2001, **19**, (10), pp. 971–974

[17] SAHA K., KEUNG A.J., IRWIN E.F., *ET AL.*: 'Substrate modulus directs neural stem cell behavior', *Biophys J.*, 2008, **95**, (9), pp. 4426–4438

[18] O'NEILL A., SCHAFFER D.V.: 'The biology and engineering of stem-cell control', *Biotechnol Appl Biochem.*, 2004, **40**, (Pt 1), pp. 5–16

[19] LAI K., ROBERTSON M.J., SCHAFFER D.V.: 'The sonic hedgehog signaling system as a bistable genetic switch', *Biophys. J.*, 2004, **86**, (5), pp. 2748–2757

[20] CHICKARMANE V., TROEIN C., NUBER U.A., SAURO H.M., PETERSON C.: 'Transcriptional dynamics of the embryonic stem cell switch', *PLoS Comput. Biol.*, 2006, **2**, (9), e123

[21] ROEDER I., GLAUCHE I.: 'Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1', *J. Theor. Biol.*, 2006, **241**, (4), pp. 852–865

[22] SHIMOJO H., OHTSUKA T., KAGEYAMA R.: 'Oscillations in notch signaling regulate maintenance of neural progenitors', *Neuron*, 2008, **58**, (1), pp. 52–64

[23] REEVES G.T., MURATOV C.B., SCHUPBACH T., SHVARTSMAN S.Y.: 'Quantitative models of developmental pattern formation', *Dev. Cell*, 2006, **11**, (3), pp. 289–300

[24] WOOLF P.J., PRUDHOMME W., DAHERON L., DALEY G.Q., LAUFFENBURGER D.A.: 'Bayesian analysis of signaling networks governing embryonic stem cell fate decisions', *Bioinformatics*, 2005, **21**, (6), pp. 741–753

[25] SHAROV A.A., PIAO Y., MATOBA R., *ET AL.*: 'Transcriptome analysis of mouse stem cells and early embryos', *PLoS Biol.*, 2003, **1**, (3), E74

[26] AIBA K., SHAROV A.A., CARTER M.G., FORONI C., VESCOVI A.L., KO M.S.: 'Defining a developmental path to neural fate by global expression profiling of mouse embryonic stem cells and adult neural stem/progenitor cells', *Stem Cells*, 2006, **24**, (4), pp. 889–895

[27] PRUDHOMME W., DALEY G.Q., ZANDSTRA P., LAUFFENBURGER D.A.: 'Multivariate proteomic analysis of murine embryonic stem cell self-renewal versus differentiation signaling', *Proc. Natl. Acad. Sci. USA*, 2004, **101**, (9), pp. 2900–2905

[28] KIROUAC D.C., ZANDSTRA P.W.: 'The systematic production of cells for cell therapies', *Cell Stem Cell*, 2008, **3**, (4), pp. 369–381

[29] MAHDAVI A., DAVEY R.E., BHOLA P., YIN T., ZANDSTRA P.W.: 'Sensitivity analysis of intracellular signaling pathway kinetics predicts targets for stem cell fate control', *PLoS Comput. Biol.*, 2007, **3**, (7), e130

[30] VISWANATHAN S., BENATAR T., ROSE-JOHN S., LAUFFENBURGER D.A., ZANDSTRA P.W.: 'Ligand/receptor signaling threshold (LIST) model accounts for gp130-mediated embryonic stem cell self-renewal responses to LIF and HIL-6', *Stem Cells*, 2002, **20**, (2), pp. 119–138

[31] ZANDSTRA P.W., LAUFFENBURGER D.A., EAVES C.J.: 'A ligand–receptor signaling threshold model of stem cell differentiation control: a biologically conserved mechanism applicable to hematopoiesis', *Blood*, 2000, **96**, (4), pp. 1215–1222

[32] BOYER L.A., LEE T.I., COLE M.F., *ET AL.*: 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell*, 2005, **122**, (6), pp. 947–956

[33] IVANOVA N., DOBRIN R., LU R., *ET AL.*: 'Dissecting self-renewal in stem cells with RNA interference', *Nature*, 2006, **442**, (7102), pp. 533–538

[34] LOH Y.H., WU Q., CHEW J.L., *ET AL.*: 'The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells', *Nat. Genet.*, 2006, **38**, (4), pp. 431–440

[35] KIM J., CHU J., SHEN X., WANG J., ORKIN S.H.: 'An extended transcriptional network for pluripotency of embryonic stem cells', *Cell*, 2008, **132**, (6), pp. 1049–1061

[36] HUANG S., GUO Y.P., MAY G., ENVER T.: 'Bifurcation dynamics in lineage-commitment in bipotent progenitor cells', *Dev. Biol.*, 2007, **305**, (2), pp. 695–713

[37] COLLIER J.R., MONK N.A., MAINI P.K., LEWIS J.H.: 'Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling', *J. Theor. Biol.*, 1996, **183**, (4), pp. 429–446

[38] WEARING H.J., OWEN M.R., SHERRATT J.A.: 'Mathematical modelling of juxtacrine patterning', *Bull. Math. Biol.*, 2000, **62**, (2), pp. 293–320

[39] MONK N.A.: 'Oscillatory expression of Hes1'p53, and NF-kappaB driven by transcriptional time delays', *Curr. Biol.*, 2003, **13**, (16), pp. 1409–1413

[40] JENSEN M.H., SNEPPEN K., TIANA G.: 'Sustained oscillations and time delays in gene expression of protein Hes1', *FEBS Lett.*, 2003, **541**, (1–3), pp. 176–177

[41] AGRAWAL S., ARCHER C., SCHAFFER D.V.: 'Computational models of the notch network elucidate mechanisms of

context-dependent signaling', *PLoS Comput. Biol.*, 2009, **5**, (5), pp. e1000390

[42] WILLERTH S.M., SAKIYAMA-ELBERT S.E.: 'Kinetic analysis of neurotrophin-3-mediated differentiation of embryonic stem cells into neurons', *Tissue Eng. A*, 2009, **15**, (2), pp. 307–315

[43] DAVEY R.E., ONISHI K., MAHDAVI A., ZANDSTRA P.W.: 'LIF-mediated control of embryonic stem cell self-renewal emerges due to an autoregulatory loop', *FASEB J.*, 2007, **21**, (9), pp. 2020–2032

[44] SIMINOVITCH L., MCCULLOCH E.A., TILL J.E.: 'The distribution of colony-forming cells among spleen colonies', *J. Cell Physiol.*, 1963, **62**, pp. 327–336

[45] TILL J.E., MCCULLOCH E.A., SIMINOVITCH L.: 'A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells', *Proc. Natl. Acad. Sci. USA*, 1964, **51**, pp. 29–36

[46] TILL J.E., MCCULLOCH E.A., SIMINOVITCH L.: 'Isolation of variant cell lines during serial transplantation of hematopoietic cells derived from fetal liver', *J. Natl. Cancer Inst.*, 1964, **33**, pp. 707–720

[47] MCADAMS H.H., ARKIN A.: 'Stochastic mechanisms in gene expression', *Proc. Natl. Acad. Sci. USA*, 1997, **94**, (3), pp. 814–819

[48] RAJ A., VAN OUDENAARDEN A.: 'Nature, nurture, or chance: stochastic gene expression and its consequences', *Cell*, 2008, **135**, (2), pp. 216–226

[49] ARKIN A., ROSS J., MCADAMS H.H.: 'Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells', *Genetics*, 1998, **149**, (4), pp. 1633–1648

[50] MSADEK T.: 'When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*', *Trends Microbiol.*, 1999, **7**, (5), pp. 201–207

[51] VILAR J.M., KUEH H.Y., BARKAI N., LEIBLER S.: 'Mechanisms of noise-resistance in genetic oscillators', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, (9), pp. 5988–5992

[52] DAS J., HO M., ZIKHERMAN J., *ET AL.*: 'Digital signaling and hysteresis characterize RAS activation in lymphoid cells', *Cell*, 2009, **136**, (2), pp. 337–351

[53] RAO C.V., WOLF D.M., ARKIN A.P.: 'Control, exploitation and tolerance of intracellular noise', *Nature*, 2002, **420**, (6912), pp. 231–237

[54] ARTYOMOV M.N., DAS J., KARDAR M., CHAKRABORTY A.K.: 'Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities', *Proc. Natl. Acad. Sci. USA*, 2007, **104**, (48), pp. 18958–18963

[55] METTETAL J.T., MUZZEY D., PEDRAZA J.M., OZBUDAK E.M., VAN OUDENAARDEN A.: 'Predicting stochastic gene expression dynamics in single cells', *Proc. Natl. Acad. Sci. USA*, 2006, **103**, (19), pp. 7304–7309

[56] GRAF T., STADTFELD M.: 'Heterogeneity of embryonic and adult stem cells', *Cell Stem Cell*, 2008, **3**, (5), pp. 480–483

[57] CHAMBERS I., SILVA J., COLBY D., *ET AL.*: 'Nanog safeguards pluripotency and mediates germline development', *Nature*, 2007, **450**, (7173), pp. 1230–1234

[58] CHANG H.H., HEMBERG M., BARAHONA M., INGBER D.E., HUANG S.: 'Transcriptome-wide noise controls lineage choice in mammalian progenitor cells', *Nature*, 2008, **453**, (7194), pp. 544–547

[59] NEEDHAM C.J., BRADFORD J.R., BULPITT A.J., WESTHEAD D.R.: 'A primer on learning in Bayesian networks for computational biology', *PLoS Comput. Biol.*, 2007, **3**, (8), e129

[60] ALTEROVITZ G., LIU J., AFKHAMI E., RAMONI M.F.: 'Bayesian methods for proteomics', *Proteomics*, 2007, **7**, (16), pp. 2843–2855

[61] SMITH V.A., JARVIS E.D., HARTEMINK A.J.: 'Influence of network topology and data collection on network inference', *Pac. Symp. Biocomput.*, 2003, **8**, pp. 164–175

[62] YU J., SMITH V.A., WANG P.P., HARTEMINK A.J., JARVIS E.D.: 'Advances to Bayesian network inference for generating causal networks from observational biological data', *Bioinformatics*, 2004, **20**, (18), pp. 3594–3603

[63] HUANG J.C., MORRIS Q.D., FREY B.J.: 'Bayesian inference of MicroRNA targets from sequence and expression data', *J. Comput. Biol.*, 2007, **14**, (5), pp. 550–563

[64] LI H., LU L., MANLY K.F., *ET AL.*: 'Inferring gene transcriptional modulatory relations: a genetical genomics approach', *Hum. Mol. Genet.*, 2005, **14**, (9), pp. 1119–1125

[65] SEBASTIANI P., RAMONI M.F., NOLAN V., BALDWIN C.T., STEINBERG M.H.: 'Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia', *Nat. Genet.*, 2005, **37**, (4), pp. 435–440

[66] JANES K.A., YAFFE M.B.: 'Data-driven modelling of signal-transduction networks', *Nat. Rev. Mol. Cell. Biol.*, 2006, **7**, (11), pp. 820–828

[67] BRIGGMAN K.L., ABARBANEL H.D., KRISTAN W.B.: 'Optical imaging of neuronal populations during decision-making', *Science*, 2005, **307**, (5711), pp. 896–901

doi: 10.1049/iet-syb.2009.0011

[68] BUTTE A.: 'The use and analysis of microarray data', *Nat. Rev. Drug Discov.*, 2002, **1**, (12), pp. 951–960

[69] TANAKA M., BATEMAN R., RAUH D., *ET AL*.: 'An unbiased cell morphology-based screen for new, biologically active small molecules', *PLoS Biol.*, 2005, **3**, (5), e128

[70] LIU G., SWIHART M.T., NEELAMEGHAM S.: 'Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling', *Bioinformatics*, 2005, **21**, (7), pp. 1194–1202

[71] JANES K.A., ALBECK J.G., GAUDET S., SORGER P.K., LAUFFENBURGER D.A., YAFFE M.B.: 'A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis', *Science*, 2005, **310**, (5754), pp. 1646–1653

[72] JANES K.A., KELLY J.R., GAUDET S., ALBECK J.G., SORGER P.K., LAUFFENBURGER D.A.: 'Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data', *J. Comput. Biol.*, 2004, **11**, (4), pp. 544–561

[73] GONCALVES A., ESTERNI B., BERTUCCI F., *ET AL*.: 'Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy', *Oncogene*, 2006, **25**, (7), pp. 981–989

[74] KUMAR N., WOLF-YADLIN A., WHITE F.M., LAUFFENBURGER D.A.: 'Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data', *PLoS Comput. Biol.*, 2007, **3**, (1), e4

[75] MILLER-JENSEN K., JANES K.A., BRUGGE J.S., LAUFFENBURGER D.A.: 'Common effector processing mediates cell-specific responses to stimuli', *Nature*, 2007, **448**, (7153), pp. 604–608