

REVIEW

Computational modelling of embryonic stem-cell fate control

Maria Herberg and Ingo Roeder*

ABSTRACT

The maintenance of pluripotency in embryonic stem cells (ESCs), its loss during lineage specification or its re-induction to generate induced pluripotent stem cells are central topics in stem cell biology. To uncover the molecular basis and the design principles of pluripotency control, a multitude of experimental, but also an increasing number of computational, studies have been published. Here, we consider recent reports that apply computational or mathematical modelling approaches to describe the regulatory processes that underlie cell fate decisions in mouse ESCs. We summarise the principles, the strengths and potentials but also the limitations of different computational strategies.

KEY WORDS: Computational biology, Embryonic stem cells, Multi-scale models, Network dynamics, Stem cell fate decisions

Introduction

Pluripotency is defined as the capacity of an individual cell to give rise to all lineages of the mature organism, including the germ line. This property is transient and restricted to a few epiblast cells in the early mammalian embryo, but it can be sustained *in vitro* through the establishment of embryonic stem cells (ESCs) (Evans and Kaufman, 1981; Martin, 1981). Mouse ESCs (mESCs) can be maintained in a naïve pluripotent state over many passages without losing their differentiation potential. Moreover, when injected into a host embryo, they return to normal development and contribute to chimera formation (Smith, 2001). Uncovering the molecular basis and understanding the design principles of pluripotency is, therefore, of utmost relevance for both stem cell research and developmental biology.

Like many other areas of modern life sciences, stem cell biology is increasingly influenced by novel technologies, such as molecular high-throughput techniques and (ultra-)high-resolution imaging. These new methods provide an unprecedented amount of information at levels previously inaccessible. Although we recognise the tremendous potential of this new level of quantification, it also generates new challenges, including new sources of error in experimental setup or analysis. Another challenge comes with the inherent complexity of biological systems: in the majority of cases, a functional and mechanistic understanding of the dynamic behaviour cannot be achieved by simply ‘adding up’ the individual effects observed for the different components of the system. As already Aristotle recognised, “the whole is more than the sum of its parts”, or, in more modern terms, the complex interplay of the individual components might lead to the emergence of unexpected behaviours.

It has become increasingly clear that a complementary application of methodologies from bioinformatics, statistics and dynamic modelling is required for an accurate description and a comprehension of complex interactions and emergent behaviours

(Fig. 1). For example, computational approaches can define the structure and the dynamics of gene regulatory networks (GRNs, see Glossary, Box 1). However, care is required when interpreting such studies, as different structures potentially encode the same or similar biological functions. Moreover, the same network structure might generate different functional consequences. To distinguish between different potential explanations, the analysis of the network dynamics is as important as structure identification. Because such network dynamics might be non-intuitive, mathematical concepts and computational methods from dynamic systems theory are extremely helpful for their formal description and quantitative analysis. The fields of application of computational models are as versatile as the methodologies themselves (see Box 2).

In this Review, we will discuss strategies to build and understand computational models of GRNs in mESCs and in the blastocyst. First, we introduce the transcriptional core network of pluripotency. Then, we describe dynamic small-scale networks probing mechanisms that underlie transcription factor (TF) heterogeneity. We also summarise a selection of more complex pluripotency networks. Finally, we consider dynamic models of cellular reprogramming and lineage specification, with a brief outlook on the demand for multi-scale modelling. For a more detailed discussion on the biological aspects of mESC pluripotency and heterogeneity, we refer the reader to other, excellent reviews (Niwa, 2007; Chambers and Tomlinson, 2009; Martinez Arias and Brickman, 2011; Nichols and Smith, 2012; Torres-Padilla and Chambers, 2014).

Dissecting the complexity of the pluripotency network through computational models

Over the last decade, microarray and sequencing technologies have been used extensively to measure gene expression in mouse and human ESCs with the aim of identifying master regulators of pluripotency regulation and lineage commitment, and their interactions on a genome-wide scale (Ivanova et al., 2006; Loh et al., 2006; Matoba et al., 2006; Wang et al., 2006; Kim et al., 2008; Ding et al., 2009; Dowell et al., 2013). Although these large-scale studies uncovered an inherent complexity of pluripotency regulation, most attention has been paid to a core unit of three TFs, namely Oct4, Sox2 and Nanog. More than 15 years ago, Oct4 was shown to be essential for the establishment of pluripotency in the embryo and in cultured cells (Nichols et al., 1998). However, although a constant expression of Oct4 is required for self-renewal, the dosage also matters: both enhanced and reduced levels of Oct4 lead to lineage specification (Niwa et al., 2000). Oct4 cooperates with Sox2 to regulate the expression of multiple target genes (Chambers and Tomlinson, 2009), including fibroblast growth factor 4 (Fgf4) that initiates lineage specification (Kunath et al., 2007; Silva and Smith, 2008). The TF Nanog supports Oct4 and Sox2 in coordinating the activity of a number of target genes. Overexpression of Nanog alleviates the requirement for extrinsic culture medium supplements, such as leukaemia inhibitor factor (LIF) or the mitogen kinase inhibitor (MEKi), to prevent differentiation (Chambers et al., 2003; Mitsui et al., 2003; Ying

Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden D-01307, Germany.

*Author for correspondence (ingo.roeder@tu-dresden.de)

Task	Examples	Scale	Method classes	Examples														
Identification of network components	Candidate genes 	Intra-cellular	Descriptive and explorative statistics	Text mining Testing for differential gene expression														
Identification of network topology	Relation of genes 			Correlation analysis Principle component analysis														
Identification of network logic	Control rules <table border="1"> <tr> <td>Gene A</td> <td>Gene B</td> <td>Gene C</td> <td>Target Gene</td> </tr> <tr> <td>ON</td> <td>ON</td> <td>ON</td> <td>ON</td> </tr> <tr> <td>ON</td> <td>OFF</td> <td>ON</td> <td>OFF</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </table>			Gene A	Gene B	Gene C	Target Gene	ON	ON	ON	ON	ON	OFF	ON	OFF
Gene A	Gene B	Gene C	Target Gene															
ON	ON	ON	ON															
ON	OFF	ON	OFF															
...															
Identification of network dynamics	Stability Patterning 	Inter-cellular	Dynamic modelling	Differential equations Cellular automata/ Potts model Multi-scale and agent-based models														

Fig. 1. General overview of computational methodologies and their applications. To reconstruct network structures and reveal their underlying dynamics at different scales, methods from statistics (e.g. correlation analysis), bioinformatics (e.g. Bayesian networks) and mathematical modelling (e.g. differential equations) can be applied. The usage of these methods is not limited to one specific task. They rather provide frameworks that can be adapted or extended to assist the exploration of many different tasks.

et al., 2008). Although Nanog is essential to reach a pluripotent cell state, pluripotency can be maintained even in the absence of Nanog, albeit in an unstable fashion (Chambers et al., 2007; Silva et al., 2009).

Dynamic properties of the pluripotency core network

Multiple experimental studies indicate that the core network constituted by Oct4, Sox2 and Nanog is characterised by a number of positive-feedback regulations (Chew et al., 2005; Rodda et al., 2005; Loh et al., 2006). The first ordinary differential equation (ODE) model (see Glossary, Box 1) of this core structure was published by Chickarmane et al. in 2006 (Fig. 2A) (Chickarmane et al., 2006). Here, the transcriptional regulation of the network components is mathematically described using Hill equations, which are widely used to model chemical reactions and approximate molecular interactions (Weiss, 1997). However, to simulate TF concentrations within this framework, parameters such as binding or transcription rates, as well as cooperativity parameters (that describe the stoichiometry of molecules) have to be defined. Unfortunately, a quantitative determination of biochemical parameters is challenging, if not impossible, and fitting them to experimental data often yields large uncertainties (Gutenkunst et al., 2007). Methodologies from dynamic systems theory, such as state space and bifurcation analysis (Box 3), are powerful tools to investigate the behaviour of a system qualitatively and quantitatively. Together with simulation studies and proper sensitivity analyses of model parameters, these approaches can identify whether the dynamics are consistent with the experimental observations and allow the effect of changes in network components or interaction rates to be predicted.

Chickarmane et al. investigated the dependency of stable Oct4 and Nanog expression states on external signals, demonstrating that the positive-feedback loops in the core network give rise to a system property termed bistability (Box 3) (Chickarmane et al., 2006). Bistability can be interpreted as the coexistence of two stable gene expression states associated with different developmental potentials of cells. The emergence and disappearance of stable expression levels as a function of a particular model parameter can be depicted by a bifurcation diagram, as shown in Fig. 2B. It reveals that the model illustrated in Fig. 2A can account for three distinct patterns of

stable expression states, depending on the intensity of an input signal B, which negatively regulates Nanog transcription (Fig. 2B). The authors nominated bone morphogenetic protein 4 (BMP4) signalling and p53 as potential candidate factors for B. When signal B is weak, the pluripotency network is sustained and a stable expression state at high Nanog levels is established (termed ‘stem-cell state’, upper solid line in Fig. 2B). By contrast, if signal B is strong, Nanog autoregulation is suppressed and its transcription remains at a basal level. Thus, a second stable state at low Nanog levels is established (termed ‘differentiated state’, lower solid line in Fig. 2B). However, for an intermediate range of signal B, both states coexist, such that mESCs can attain either of the two expression states (region shaded in grey in Fig. 2B). Thus, a continuous increase of signal B intensity ‘moves’ cells from a monostable high into a monostable low state by transiently passing the bistable region. This process is reversible but the signal intensities at which cells change their state into one or the other direction (c_1 and c_2) are different. This property is called hysteresis and means that the behaviour of a system depends on its history. Due to the highly connected architecture of the underlying network, the model predicts the same qualitative behaviour for Oct4 and Sox2. Thus, in the model proposed by Chickarmane et al., all three TFs are either on or off, ensuring that the expression of common target stem cell and differentiation genes is mutually exclusive.

As shown above, the analysis of qualitative changes of gene expression patterns depending on model parameters provides insights into potential cellular states, their stability and robustness. However, to study dynamic properties of developing cell populations, the heterogeneity of gene expression and of cellular states has to be taken into account.

The role of gene expression heterogeneity

The maintenance of mESC pluripotency requires specific culture conditions. Standard mESC culture media contain serum factors and the cytokine LIF (Niwa et al., 1998), and promote a population-intrinsic heterogeneity reflected in the mosaic expression of pluripotency genes, including *Nanog*, *Rex1* (also known as *Zfp42* – Mouse Genome Informatics Database), *Stella* (also known as *Dppa3* – Mouse Genome Informatics Database) and *Esrrb*

Box 1. Glossary

Bayesian network model: Probabilistic model that describes the interactions of network components (random variables) in terms of their conditional dependencies (Pe'er, 2005).

Boolean network model: Qualitative model in which network components are represented by Boolean variables. The state of components is determined as a logical function of the states of all the components linked to it (Bornholdt, 2008).

Clustering: Collection of algorithms that organises datapoints into groups according to some measures of similarity. The selection of these measures sensitively determines the final result (Frades and Matthiesen, 2010).

Correlation analysis: Statistical technique that uses correlation coefficients to describe stochastic dependence. Statistical correlation must not be confused with causal relation (Bewick et al., 2003).

Dynamic systems theory: Provides a mathematical framework to describe the temporal behaviour of complex systems. Specifically, it focuses on stability properties, i.e. analysing which system states can be expected to be stable over time and under which conditions.

Gene regulatory network (GRN): The structure (i.e. the topology) of a GRN is defined by a number of components (e.g. genes) and their regulatory interactions. It can be represented by a directed graph consisting of nodes (components) and edges (interactions). The network dynamics (i.e. temporal changes of component states) is not purely determined by the topology, but also depends on the parameters describing the individual interactions and the initial conditions.

K-means clustering: Clustering algorithm that partitions a dataset into a number of k pre-specified clusters by minimising the sum of squared Euclidian distances of individual datapoints from the cluster centroids.

Multi-scale model: Mathematical framework that combines the description of different temporal and/or spatial scales. Multi-scale models are able to explain behaviours that emerge on one scale, on the basis of effects that are acting on other scales (Nakaoka, 2014).

Ordinary differential equations (ODE): Mathematical formalism to describe continuous changes of variables (e.g. proteins) using a deterministic function of the variable and its derivatives.

Pearson's correlation coefficient: Measure of the direction and strength of a linear relationship between two variables (Taylor, 1990).

Principal component analysis (PCA): Statistical method that describes given datapoints in terms of linearly uncorrelated variables, denoted as principal components (PC), leading to a dimensionality reduction (Ringnér, 2008). In general, PCs are constructed in such a way that the first PC explains the largest part of the variance in the dataset, followed by the second, third, etc.

Stochastic differential equations (SDE): Differential equations that describe stochastic changes of variables (e.g. proteins) by incorporating a random noise term. As analytic solutions are rarely possible, they are often analysed by numerical algorithms or by simulations (e.g. Monte-Carlo simulation).

Support vector machine (SVM): Procedure to classify datapoints into pre-defined categories. SVMs are examples of supervised learning algorithms, because it uses training data with known categorisation (Noble, 2006).

Transcriptional noise: Fluctuations in gene expression can be modelled by incorporating a stochastic term to the differential equations (SDEs). Additive noise: Noise amplitude is independent of gene expression levels. Multiplicative noise: Noise amplitude increases linearly with expression levels.

(Chambers et al., 2007; Carter et al., 2008; Toyooka et al., 2008). Because of the key role of Nanog in the establishment of pluripotency, its heterogeneity has been studied extensively. Flow cytometry analyses of Nanog reporter cell lines show a stable and reproducible bimodal distribution of Nanog-high (NH) and Nanog-low (NL) cells (Chambers et al., 2007; Kalmar et al., 2009). Purified NH cells can give rise to NL cells and vice versa, revealing the dynamic capacity of mESCs to undergo state transitions (Box 3),

Box 2. Fields of application of computational models

The development and the refinement of a computational model is an iterative process of (i) formulating assumptions (based on experimental observations), (ii) implementing these assumptions in the context of a mathematical description, (iii) evaluating the model (e.g. by computer simulations) and (iv) comparing model results and predictions with experimental data. The application of models aims at revealing insights into the mechanisms underlying the properties or the regulatory processes of a system.

In particular, computational models

- summarise our understanding of a system or a process and, thus, help to identify gaps in our knowledge;
- provide presumptive mechanisms to explain and link a variety of observed phenomena and reveal to which extent data are consistent between experiments and with theoretical concepts;
- identify the general principles that govern a biological system;
- allow to formulate hypotheses about the qualitative and the quantitative behaviour of a system;
- can be used to perform experiments *in silico* that would be too elaborative or costly to be conducted at the bench;
- anticipate the impact of the manipulation of a system or its perturbation;
- allow us to analyse, visualise and interpret complex datasets.

To be of practical relevance, computational models have to

- be based on clearly defined and biologically plausible assumptions;
- be described comprehensibly;
- provide experimentally testable predictions.

and show a functional bias in their differentiation propensity (Singh et al., 2007; Marks et al., 2012). The biological relevance (Smith, 2013) and the molecular mechanisms underlying this intrinsic heterogeneity are still under debate (Miyazari and Torres-Padilla, 2012; Navarro et al., 2012; Karwacki-Neisius et al., 2013; Muñoz-Descalzo et al., 2013; Singer et al., 2014).

Complementary to experimental strategies, a few small-scale GRN models have been applied to improve the mechanistic understanding of dynamic heterogeneity in pluripotent mESCs (Kalmar et al., 2009; Glauche et al., 2010; Herberg et al., 2014). Three different mechanisms accounting for a bimodal Nanog distribution and reversible state transitions will be discussed here: first, an excitable model presuming monostable Nanog expression and extrinsic noise (Kalmar et al., 2009); second, an oscillation model based on a cyclic Nanog attractor (Glauche et al., 2010); and third, a fluctuation model assuming bistable Nanog expression and intrinsic noise (Glauche et al., 2010; Herberg et al., 2014). We review these models to evaluate their consistency with new experimental findings based on single-cell gene expression recently published by Singer et al. (2014).

In 2009, Kalmar et al. proposed a noise-driven, excitable model based on positive autoregulations and mutual activations of Oct4 and Nanog (Fig. 3A) (Kalmar et al., 2009). Additionally, the authors supposed that sufficiently high levels of Oct4 can repress Nanog expression, introducing a negative feedback. To formally describe these regulations, a stochastic differential equation (SDE) model (see Glossary, Box 1) has been used. Although the positive-feedback loops allow for bistability, parameters have been chosen such that only one stable NH state exists (compare with Fig. 2B: 'Stem cell state'). Excursions from this state have been induced by a stochastic noise term (see 'additive noise' under the entry 'Transcriptional noise' in the Glossary, Box 1). Thus, mESCs stochastically escape from the stable

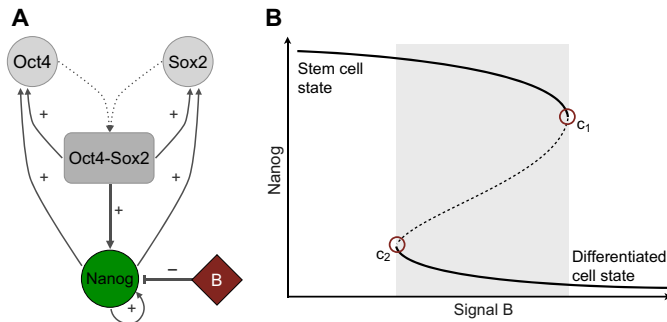


Fig. 2. Core network of pluripotency regulation. (A) Scheme of the network structure suggested by Chickarmane et al. (2006). Oct4 and Sox2 form a heterodimer (dotted lines) that activates Oct4, Sox2 and Nanog transcription. Nanog regulates its own transcription and activates Oct4 and Sox2. An external signal, termed B, suppresses Nanog expression. (B) Corresponding bifurcation diagram of Nanog expression as a function of the levels of signal B. Solid lines represent stable states, dashed lines unstable states. The bistable region is shaded in grey. Bifurcations occur at signal intensities c_1 and c_2 .

expression state, which is characterised by high Oct4 and Nanog expression, towards a region of low Nanog and variable Oct4 expression. As no second (attracting) stable state exists, cells rapidly return to their origin and, thus, exhibit pulsing Nanog dynamics (Fig. 3A, bottom). This model predicted that excursions from the NH state are transient, providing a very short window of opportunity in which perturbations can become consolidated into a lineage commitment decision (Chambers et al., 2007).

Exploring the same topic, Glauche et al. suggested two alternative SDE models (Glauche et al., 2010). In the first model (model 1 in Fig. 3B) mESCs are bistable with respect to Nanog expression (compare with Fig. 2B, grey shaded region). State transitions are

induced by a transcriptional noise (see ‘multiplicative noise’ under the entry ‘Transcriptional noise’ in the Glossary, Box 1) leading to stochastic fluctuations. In the second model (model 2 in Fig. 3B), a hypothetical factor X has been integrated such that Nanog is part of an activator-repressor system. The resulting negative feedback can generate a cyclic attractor and thus oscillations between high and low Nanog expression levels (Fig. 3B, bottom). Replacing the previously suggested direct activation of Oct4-Sox2 by Nanog with an indirect, double-negative feedback, Glauche et al. suggested a mechanistic explanation for the gatekeeper function of Nanog in the control of mESC differentiation (Silva and Smith, 2008). Indeed, assuming that Nanog prevents the effects of differentiation signals (represented by Y), only mESCs in the NL state can differentiate. Meanwhile, Muñoz-Descalzo et al. provided an alternative explanation, showing that Nanog buffers the differentiation-inducing activity of Oct4 through the formation of stable complexes (Muñoz-Descalzo et al., 2013) (see section below on lineage specification).

Recently, Herberg et al. put forward the fluctuation model to study whether different intensities of Fgf4/Erk signalling can account for culture-dependent differences in Nanog expression (Herberg et al., 2014). mESCs exposed to serum-free medium containing MEK/Erk and Gsk3 β inhibitors, termed 2i, are captured in a pluripotent ground state characterised by a single, homogenous NH peak (Ying et al., 2008; Wray et al., 2010). Applying the model illustrated in Fig. 3C, the authors demonstrated that Fgf4/Erk signalling can regulate the existence of stable Nanog expression patterns by ‘shifting’ cells from a monostable high into a bistable regime as described above (compare with Fig. 2B, with signal B assumed to represent Fgf4/Erk). In particular, if Fgf4/Erk signalling is efficiently blocked (as in 2i), there is only one stable expression state at high Nanog levels. In this ground state, network-inherent stochasticity or (external) perturbations only induce micro-heterogeneity within an apparently uniform cell population. However, if Fgf4/Erk signalling is active, two stable expression states coexist (bistability), such that stochastic fluctuations generate functionally different subpopulations (macro-heterogeneity, Fig. 3C, bottom) (Huang, 2009; Herberg et al., 2014).

Although all three models consistently describe a stable, albeit dynamic, bimodal distribution of Nanog expression within a population of mESCs (compare with insets in Fig. 3, bottom row), they generate different hypotheses about the underlying mechanism and its dynamic outcome. These hypotheses guided experimental strategies to distinguish between the mechanisms (Box 2). For example, Glauche et al. demonstrated that stochastic fluctuations and deterministic oscillations are distinguishable based on the dynamics of the (re-)establishment of heterogeneity after cell sorting (Glauche et al., 2010). Furthermore, comparing the models revealed that measurements of the residence times of single cells in the NH and the NL state would allow the discrimination of mono- from bistable fluctuations. In a bistable system, residence times for both Nanog expression states would be prolonged, exceeding typical cell cycle times as predicted by Herberg et al. (2014), whereas a monostable system implies that unstable NL cells rapidly revert back into the stable NH state (Kalmar et al., 2009). Intriguingly, Singer et al. recently analysed single-cell expression dynamics, demonstrating that Nanog heterogeneity arises from a combination of stochastic transitions between coexisting gene expression states (i.e. macro-heterogeneity, bistability) and burst-like transcription within each state (i.e. micro-heterogeneity, noise) (Singer et al., 2014). They also showed that single mESCs remain in either one of the two Nanog states for multiple cell cycles. These findings clearly favour the fluctuation model suggested and described by Glauche et al. (2010) and Herberg et al. (2014).

Box 3. The state space formalism: a model to quantitatively describe cell fate decisions

The gene expression state S of a cell is defined by the expression values of a set of genes at a certain time point. This state is a point in an abstract space composed of all possible expression states, termed state space. As gene expression changes due to stochastic fluctuations or developmental processes, the position of state S in the state space changes. The particular path that is taken is denoted as a trajectory. Because of regulatory interactions between genes, not all expression states are actually reachable.

Differential equations can be used to describe mathematically how gene expression evolves over time. Thus, the solutions of a system of equations describe corresponding trajectories. States, to which trajectories move over time, are termed stable fixed points (or steady states), because once these points have been reached, the state does not change anymore without external stimulation. Unstable fixed points, by contrast, are transient, because infinitesimally small perturbations lead to a movement away from these points. Alterations in gene regulation can change stable expression patterns (fixed points) quantitatively and qualitatively. Qualitative changes, termed bifurcations, occur when fixed points emerge, disappear or change their stability.

If two or more stable fixed points coexist at a given time, the system is termed bistable or multistable, respectively. In a deterministic situation, the initial conditions, i.e. the starting point of the trajectory, determine to which stable state the trajectory moves. Changes between stable states are termed state transitions and are associated with changes in the fate or the developmental potential of a cell. State transitions can be continuous or, given a multi-stable system, switch-like in response to external perturbations or stochastic fluctuations.

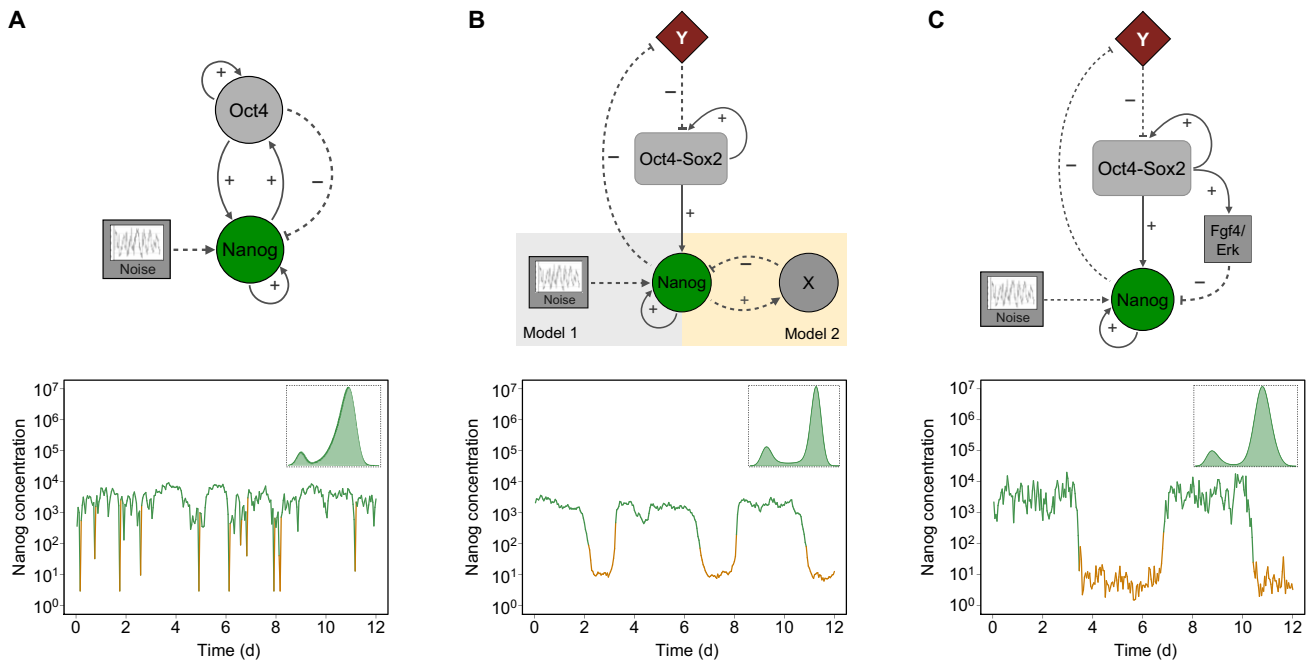


Fig. 3. Modelling small-scale networks of Nanog heterogeneity. (A) The GRN proposed by Kalmar et al. (2009) generates sudden excursions from the Nanog high (NH) state (green lines, lower panel) to the region of low expression of Nanog (orange lines, lower panel) as shown by the simulated time course. Because there is no second stable expression state, mESCs rapidly return back into the NH state. (B) GRN proposed by Glauche et al. (2010). The intracellular activity of a signalling cascade (termed Y), which inhibits the expression of the pluripotency factors, depends on the Nanog concentration of the cell. The oscillatory behaviour of model 2 (orange square, X is a hypothetical Nanog repressor) is depicted in the time course below. The outcome of model 1 (grey square) corresponds to the dynamic shown in C. (C) The GRN proposed by Herberg et al. (2014) causes stochastic fluctuations between a stable NH state and a stable NL state as shown in the time course. Residence times in the NL region are much longer compared with the model dynamics shown in A. Positive regulations in GRNs are denoted by '+', negative by '-'. Hypothetical interactions are illustrated by dashed lines. Small insets show simulated Nanog distributions at day 12.

Taken together, the complementary use of computational and experimental strategies clearly indicates that Nanog heterogeneity is based on two coexisting expression states and stochastic fluctuations at the transcriptional level. Specifically, the presented mathematical models brought up hypotheses on potential regulatory mechanisms, evaluated their consistency with available experimental data and revealed experimental strategies required to distinguish between them (Box 2).

Pluripotency regulation in the context of more complex networks

Although the GRN of mESC pluripotency is centred on Oct4, Sox2 and Nanog, numerous studies point to the involvement of additional transcriptional (co-) factors, such as Klf4, Esrrb, Dpp1, Zfp281, Tcf1 and Tbx3 (Ivanova et al., 2006; Matoba et al., 2006; Wang et al., 2006; Kim et al., 2008), to mention only a few. In this section, we introduce three studies in which statistical and bioinformatics methods have been applied to analyse more complex TF interactions under different conditions.

MacArthur et al. studied the role of Nanog heterogeneity in early fate decisions by mimicking its temporal fluctuations (MacArthur et al., 2012). In particular, they analysed global gene expression changes during downregulation of Nanog, with a focus on elements of a previously published GRN (Kim et al., 2008). All factors, including Oct4, Sox2 and Klf4, were downregulated considerably later than Nanog, providing a time window in which cells can be rescued from differentiation by reintroducing Nanog expression. Using clustering and support vector machine approaches (see Glossary, Box 1) for the analyses of single-cell expression profiles, the authors strengthened the view that mESCs exit pluripotency gradually and stochastically. They also established two classifiers that distinguish pluripotent and non-pluripotent cells by learning the

expression profiles of cell samples taken from the Gene Expression Omnibus (GEO) database (Edgar et al., 2002). The application of these classifiers showed an intermediate, primed cell state in which differentiation and pluripotency genes are co-expressed. Finally, MacArthur et al. analysed the feedback structure of the GRN and found that fluctuations in Nanog expression transiently activate different sub-networks driving transitions between a self-sustaining state (feedback-rich) and a differentiation-sensitive state (feedback-sparse). Similar results have been reported by Trott et al., who analysed single-cell gene expression levels with the objective to decipher the activity of (sub-)networks in heterogeneous mESCs (Trott et al., 2012).

Recently, Xu et al. constructed a GRN consisting of 15 pluripotency regulators and 15 lineage markers from the ESCAPE database (Xu et al., 2013, 2014). Based on single-cell gene expression data of mESCs cultured in LIF/serum and 2i/LIF conditions, the authors attempted to reveal differences in the network structure between the two conditions using a Boolean network model (see Glossary, Box 1) to learn the underlying network logic. In that model, continuous gene expression values had to be converted to binary values (0: factor is off, 1: factor is on), which was carried out using a k-means clustering step (see Glossary, Box 1). To derive the logical function of each network node, an exhaustive search with limitations on the operators to constrain the search space (e.g. nesting, XOR gates and autoregulatory interactions were neglected) was performed, yielding a group of network models consistent with the experimental data. Xu et al. used this set of models to predict several characteristics, including the most common feedback loops, the differences between network interactions in the two culture conditions and previously uncharacterised interactions. For example, the models indicate that

the positive feedback between Oct4 and Sox2 is absent in 2i/LIF conditions, but present in LIF/serum. However, this and other predictions have yet to be validated experimentally.

In a methodologically similar study, Dunn et al. applied a data-constrained Boolean model to define a minimal GRN that allowed them to reproduce expression profiles of mESC populations in different culture conditions (Dunn et al., 2014). The authors started their analysis with a model composed of 17 pre-selected factors implicated in mESC maintenance. To reveal possible interactions, expression correlations between all pairs of TFs were quantified using Pearson's correlation coefficient (see Glossary, Box 1). Subsequently, a software tool was used to constrain the set of possible models to those able to reproduce the experimental data. Dunn et al. only kept the components and interactions that are highly correlated and present in all network structures. This selection resulted in 12 components and 11 interactions in addition to pre-specified target interactions. The authors applied the candidate models to investigate the consequences of different TF knockouts and tested their predictions experimentally. Because none of the models could explain the outcome of the Tbx3/Klf2 double-knockdown, the observed behaviour after knockdown was added as a new data constraint. Similar to the results of Xu et al. (2014), the positive link between Oct4 and Sox2 did not show up during the selection process but was integrated manually to fulfil all predefined constraints and predictions in 2i/LIF conditions. The authors claimed that the constructed GRN model constituted 'the most parsimonious network' that maintains naïve pluripotency. Although combining a Boolean network model with prior information on regulatory interactions is generally feasible, in our opinion this statement should be complemented with the limiting clause 'within the given set of assumptions'.

The presented studies are examples of computational methods suitable to evaluate gene expression data and to infer (TF- or culture-dependent) interactions between larger sets of genes. However, they also make clear that modelling the dynamics of more complex GRNs requires an even more simplified description of the system. Boolean model frameworks are commonly used to describe GRNs, but they present limitations (e.g. discretised expression levels) that have to be considered when interpreting the results (see Discussion). Applied in a proper way, these models are extremely useful to predict the consequences of perturbations (Box 2) and to identify candidate genes that play decisive roles in the loss or regain of pluripotency, as discussed in the next section.

Computational models of cellular reprogramming

In 2006, Takahashi and Yamanaka evaluated a number of candidate genes with respect to their potential to induce pluripotency in somatic cells (Takahashi and Yamanaka, 2006). The discovery that the overexpression of four factors (Oct4, Sox2, Klf4 and c-Myc) can direct the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs) constituted a paradigm shift in developmental biology. However, details and mechanisms of the reprogramming process during iPSC generation still need to be fully elucidated (Takahashi and Yamanaka, 2013).

A current limitation of the cellular reprogramming process is its low efficiency, for which there are two conceptual explanatory models: the deterministic model, in which only some cells have the potential to generate iPSCs within a fixed and uniform time period (latency), and the stochastic model, in which most or even all cells are competent for reprogramming, but the latency differs (Hanna et al., 2009; Yamanaka, 2009). Hanna et al. combined time-series experiments with computational modelling to gain insights into the

nature of reprogramming (Hanna et al., 2009). In their experiments, they monitored the proportions of iPSCs generated and found that reprogrammed cells appear in most clonal populations, if cultured for a sufficiently long time. However, the time to conversion apparently differs between clones, supporting the view that reprogramming is a continuous, stochastic process. Accordingly, the authors developed a mathematical model that considers reprogramming as a one-step stochastic process with a constant cell-intrinsic rate. Fitting the model to their experimental data, this intrinsic rate was estimated for different experimental settings. The authors argued that Nanog overexpression accelerates the reprogramming kinetics through a mechanism most likely independent of the cell proliferation rate. However, in that study, the molecular changes during reprogramming were analysed in heterogeneous cell populations, making sequential events occurring in single cells inaccessible. Combining single-cell expression analysis with a Bayesian network model (see Glossary, Box 1), Buganim et al. demonstrated that the reprogramming process can best be described by two phases: an early stochastic phase with high variation in gene expression and a subsequent, more hierarchical phase of gene activation (Buganim et al., 2012). The network model predicted that the activation of Sox2 initiates a number of consecutive steps that finally lead to fully reprogrammed iPSCs. The authors applied the hierarchical model to predict sets of TFs capable to induce pluripotency. Subsequently, these sets were tested experimentally, demonstrating that all of them facilitate iPSC generation but with different efficiencies, ranging from 0.2% (a combination without Oct4, Klf4, c-Myc) to 22.2% (using Oct4, Nanog, Esrrb, Klf4 and c-Myc). Interestingly, there is limited correlation between those genes that can facilitate efficient reprogramming and those of which the expression is predictive of future iPSC generation. For example, the endogenous expression of Oct4 does not necessarily predict iPSC generation, whereas the expression of Utl1 and Esrrb does.

Given the importance of both the genetic and epigenetic control of pluripotency, Artyomov et al. developed a computational model of pluripotency induction that couples both layers of regulation (Artyomov et al., 2010). In this study, all genes responsible for a particular cellular identity (e.g. Oct4, Sox2 and Nanog for pluripotency) were described as a single ensemble module. Moreover, cellular states, arranged in a hierarchical tree-like structure, were described not only by the expression levels of master genes, but also by the state of their epigenome. Artyomov et al. defined a set of rules on how protein expression can modify the epigenetic state and vice versa during cell cycle progression. Simulating thousands of independent reprogramming experiments, the authors found rare pathways leading to successful reprogramming.

As briefly illustrated by the studies discussed above, a combination of dynamic single-cell expression data and computational models provides insights into the different phases of the reprogramming process. Although an optimisation of the process seems hardly possible within the stochastic phase, targeted activation of pathways or genes in the hierarchical phase can enhance the generation of fully reprogrammed iPSCs. Selecting cells or cell colonies that express predictive markers might help to further increase the proportion of iPSCs, especially because environmental cues and cell-cell communication, which can be manipulated experimentally, play pivotal roles (compare with section on spatio-temporal dynamics).

Whereas stimulating particular components of the pluripotency GRN is crucial for the re-acquisition of a pluripotent cell state, limiting its activity is essential for its maintenance, as shown by

studies on mESC differentiation, which we will cover in the following section.

Towards a quantitative description of cell differentiation dynamics

Mimicking the first lineage specification observed in the embryo, mESCs can differentiate into epiblast cells (Epi; giving rise to embryo proper), primitive endoderm (PrE; generating extra-embryonic membranes) and trophoblast (TE; giving rise to extra-embryonic tissues), depending on the culture environment.

The differentiation into PrE and TE can be forced by ectopic expression of the lineage determinants Gata6 for PrE (Fujikura et al., 2002) and Cdx2 for TE (Beck et al., 1995; Strumpf et al., 2005), but also through changes in the expression of Oct4. Whereas an acute repression of Oct4 leads to differentiation into TE, an increase causes differentiation into PrE (Niwa et al., 2000). Therefore, Oct4 exerts a dose-dependent dual function as both a key factor for pluripotency and a signal driving cells toward lineage specification. However, how these dual functions of Oct4 are regulated remains to be elucidated.

To describe the dose-dependent function of Oct4, Chickarmane and Peterson extended their bistable core network (ODE model; see above) by integrating two mutual antagonisms: one between Oct4 and Cdx2, and one between Nanog and Gata6 (Chickarmane and Peterson, 2008). The simplified GRN, outlined in Fig. 4A, is based on a set of experimental findings reviewed by Niwa (2007). Additionally, the authors suggested a novel motif to account for the non-linear ('bell-shaped') relation between stable Nanog states and expression of Oct4, as illustrated in Fig. 4B and described by Matoba et al. (2006). In this motif, Oct4 activates Gata6, and both factors cooperatively suppress Nanog transcription (dashed lines in Fig. 4A). With this assumption, the network model can give rise to three different dynamics, depending on Oct4 expression, which is assumed to be controlled by an external signal A (e.g. Bmp or Wnt/ β -catenin or signalling). If Oct4 expression is low, the stem cell switch including Nanog is off and Cdx2 is constantly expressed at high levels (TE-like state; Fig. 4B). Enhanced Oct4 expression initially leads to an increase of Nanog and, thus, to a stable stem cell state. However, at a certain threshold, Oct4 induces a low-level expression of Gata6 and both factors cooperatively repress Nanog transcription, leading to bistability (Fig. 4B, region shaded in grey). In this range, the fate of a cell is determined by the initial concentrations of the pluripotency factors. If Oct4 is further increased or overexpressed, the autoregulatory capacity of Gata6 is switched on and Nanog transcription is efficiently suppressed (PrE-like state; Fig. 4B,C). According to the model prediction that

the stem cell state is most 'stable' for an intermediate range of Oct4, cellular reprogramming has also been suggested to be most efficient when Oct4 is 'overexpressed' within a specific range (Chickarmane et al., 2012). Otherwise, Oct4 expression is either too low to reactivate the pluripotency network or sufficiently high to activate Gata6, which, in both cases, leads to a stable differentiation state. Consequently, the question arises as to how the concentration range of Oct4 is or can be narrowed to efficiently promote pluripotency.

Combining quantitative single-cell analysis and mathematical modelling, Muñoz-Descalzo et al. pointed out that post-translational interactions allow the buffering of the differentiation-inducing activity of Oct4 (Muñoz-Descalzo et al., 2013). They demonstrated that, under self-renewing conditions, Nanog and β -catenin proteins form stable complexes with Oct4, and, thus, restrict its levels to a range that facilitates pluripotency. However, under LIF/serum conditions, Oct4 levels are high enough to cause Nanog bistability, potentially due to the activation of Fgf4/Erk or Gata6. According to the authors, Nanog fluctuations impact the amount of unbound ('active') Oct4 proteins that trigger differentiation. In line with this, lowering Oct4 levels in pluripotent mESCs has been shown to reduce Nanog heterogeneity and to establish a more robust pluripotency state (Karwacki-Neisius et al., 2013).

Modelling the mutual transcriptional and post-transcriptional regulation of pluripotency and lineage factors provided possible mechanistic explanations for the dose-dependent activity of Oct4 and its association with the instability of the pluripotent cell state when Nanog expression is low.

From temporal to spatio-temporal dynamics

So far, we have focused on the analysis of temporal dynamics of intracellular GRNs. However, mESCs do not exist in isolation. Similar to their *in vivo* counterparts, they self-organise as spheroid aggregates, termed embryoid bodies (EBs), or arrange in colony structures. Embedded in these structures, cells send and receive signals which might affect gene expression and, thus, establish spatial patterns or cell type arrangements.

Spatio-temporal patterns in embryonic development

One example of spatio-temporal patterning can be found in the early phases of mouse development when the inner cell mass (ICM) segregates into Nanog⁺ epiblast cells (Epi) and Gata6⁺ PrE cells [reviewed by Artus and Hadjantonakis (2012)]. After an initial phase of co-expression, both TFs display a mutually exclusive salt-and-pepper pattern (Chazaud et al., 2006) and finally arrange into two layers, with PrE cells occupying the outer layer that faces the blastocoel (Rossant and Tam, 2009). This patterning depends on

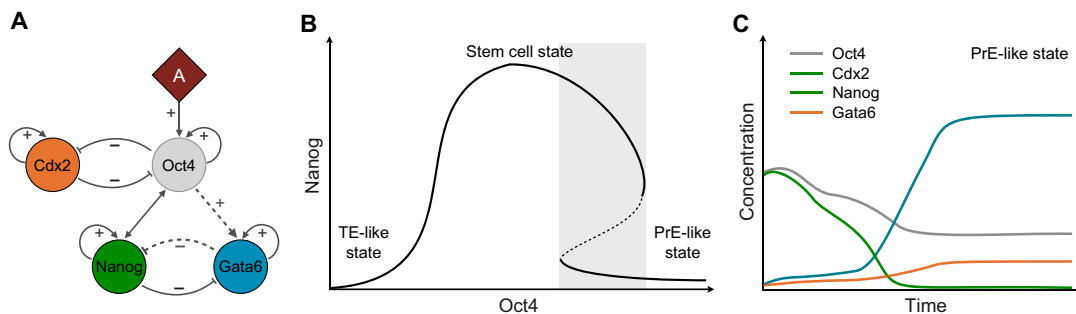


Fig. 4. Mutual antagonisms can direct lineage specification *in vitro*. (A) Scheme of the GRN suggested by Chickarmane and Peterson (2008). The interactions between Oct4, Gata6 and Nanog lead to the 'bell-shaped' activation of Nanog shown in B. An external signal, termed A, positively regulates Oct4 expression. (B) Schematic bifurcation diagram of Nanog states as a function of Oct4. Solid lines represent stable states, dashed lines unstable states. The bistable region is shaded in grey. (C) Schematic time course for PrE differentiation adapted from Chickarmane and Peterson (2008).

Fgf4 signalling, which is induced by cells that express Nanog (Yamanaka et al., 2010; Frankenberg et al., 2011).

To reveal a potential mechanism for the establishment of the salt-and-pepper pattern, Bessonnard et al. developed an ODE model describing the interactions between Nanog, Gata6 and the Fgf4/Erk signalling (Fig. 5A) (Bessonnard et al., 2014). First, they demonstrated that the model can account for the coexistence of three stable expression states within a small range of Fgf4 activity (Fig. 5B). These stable states correspond to an ICM-like state (co-expression of Nanog and Gata6), an Epi-like state (Nanog expression) and a PrE-like cell state (Gata6 expression, Fig. 5C). Subsequently, 25 ‘cells’ (i.e. GRNs) were arranged *in silico* on a squared grid to specify neighbour relationships between individual cells. Cell-cell communication was modelled by averaging the concentrations of Fgf4 produced by a cell itself and by its closest neighbours. This average concentration determined the intracellular activity of Erk signalling. The model assumes an initial heterogeneous distribution of Fgf4, and this leads to a random Nanog versus Gata6 expression pattern, consistent with experimental findings (Chazaud et al., 2006; Plusa et al., 2008). Thus, the model provides one potential mechanistic explanation for the salt-and-pepper cell pattern observed in the ICM. However, it involves a rather large number of arbitrary parameters that had to be confined to a specific range in order to capture the presented dynamics (i.e. tristability) (the so-called ‘over-fitting problem’, discussed further below). Another issue comes with the static nature of the spatial dimension. The ICM is subject to dynamic changes and re-arrangements (e.g. due to cell proliferation or mechanical forces), which impact the emergent pattern. However, these changes have been neglected in this study. Krupinski et al. considered these dynamic properties and combined a three-dimensional (3D) cell-based model with GRNs to simulate both the spatial structure and gene expression levels of a growing blastocyst (Krupinski et al., 2011). Herein, each cell is represented by an incompressible ellipsoid that interacts with its local environment through elastic adhesion forces. Furthermore, each cell is characterised by TF concentrations [determined according to the GRN developed by Chickarmane and Peterson (2008), described above], cell cycle and polarity. By coupling the intrinsic properties of ICM cells with the mechanical forces present in their environment, this model framework allows us to address general, but very important, questions, such as: Does cell position determine gene expression? Do geometric constraints affect spatial expression patterns? Are differential adhesion strengths sufficient to facilitate cell sorting? The ability to assess these questions *in silico* is particularly significant, as this type of question is still challenging to address

experimentally although technologies for *in vivo* imaging and quantification are constantly improving. This is partly due to the inaccessibility of emerging embryonic structures but also due to the lack of quantitative measures of single-cell properties and spatial patterns.

Multi-scale modelling reveals functional links between GRNs and cellular properties

Recent findings on the impact of cell density and clustering on the self-renewal and differentiation of mESCs (Peerani et al., 2009; van den Brink et al., 2014; Warmflash et al., 2014) emphasise once more that gene expression levels, cellular properties and spatial effects have to be considered in order to achieve a true systemic understanding of stem cell fate decisions. To link these different regulatory layers and to comprehensively analyse the emergent dynamic behaviour, multi-scale mathematical modelling is needed.

At present, live-cell imaging of fluorescently labelled cells is the most suitable method to acquire structural and transcriptional information of cells within an ‘undisturbed’ environment. However, not all cell properties influencing inter- and intracellular regulatory processes, e.g. cell adhesions or mechanics, can be measured or derived directly from imaging data. In this case, multi-scale models can be used to explore systematically the consequences that functional changes at the cellular level might have at the population level. For example, the multi-scale blastocyst-model of Krupinski et al. demonstrates how gene expression, cell proliferation and mechanical properties jointly structure an early embryo. In particular, it revealed that the formation of TE is most robust when assuming a position-based regulation of Cdx2 expression, and that differential adhesion strengths are crucial for the sorting of Epi and PrE cells (Krupinski et al., 2011). Similar relationships between cell position and TF expression have been reported by Herberg et al. (2015). Comparing spatial gene expression patterns of *in vitro* and *in silico* mESC colonies on the basis of quantitative measures revealed that cells with a high self-renewing capacity are located in the interior of a colony structure due to TF-related differences in proliferation and adhesion (Herberg et al., 2015). By simulating spatio-temporal patterns of Oct4 expression in EBs, White et al. showed that competing influences between Oct4⁺ and Oct4⁻ neighbours can account for experimentally observed clusters during pluripotency loss, independently of EB structure, size or cell division (White et al., 2013).

Although there are currently only very few multi-scale models of mESC differentiation available, their insights already indicate the importance of considering not only transcriptional but also cellular interactions in the regulation of pluripotency.

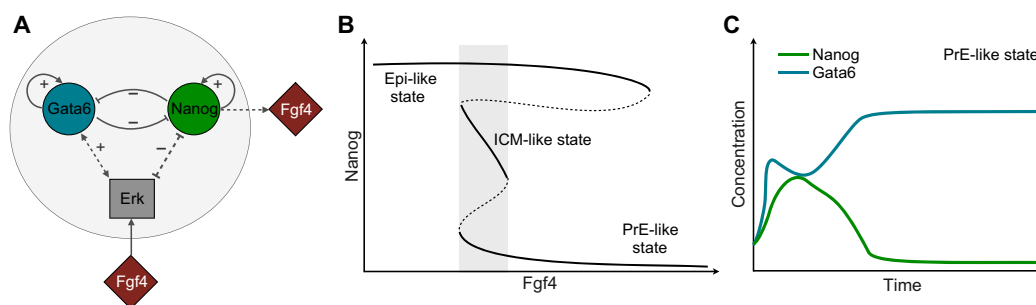


Fig. 5. Cell-cell interactions via Fgf4 can direct lineage specification *in vivo*. (A) Scheme of the GRN suggested by Bessonnard et al. (2014). (B) Schematic bifurcation diagram of Nanog states as a function of Fgf4. Solid lines represent stable states, dashed lines unstable states. The tristable region is shaded in grey. (C) Schematic time course for PrE differentiation adapted from Bessonnard et al. (2014).

Conclusion and perspectives

Here, we have discussed a number of different studies, which all apply computational approaches to quantitatively analyse regulatory processes that underlie fate decisions in mESCs and in the blastocyst. Although all these *in silico* approaches share this aim, the specific questions they aim to answer and the methods applied to do so are quite different. From a general perspective, one can categorise these questions into those aiming to establish a quantitative description of a biological process (i.e. asking “What is going on?”) and those geared towards understanding the underlying mechanisms (i.e. asking “Why does it happen?”) (Mainzer, 2014). Answering the first type of questions calls for a detailed description of the behaviour of the system, without necessarily requiring a deep understanding of the underlying mechanisms. By contrast, addressing the mechanisms that determine the behaviour of the system requires the integration of different levels of information, potentially at the expense of a detailed description of the system.

Although there are methodological overlaps, these two perspectives are usually linked to specific computational approaches (Fig. 6). Typical tools to address ‘What?’-questions are bioinformatics methodologies. The availability of public databases such as GEO (Edgar et al., 2002), FunGenES (Schulz et al., 2009) or ESCAPE (Xu et al., 2013), together with the application of sufficient computing resources, allows us to very quickly answer scientific questions such as: Which genes are differentially expressed across tissue samples? What is the effect of activating or repressing a gene in a defined context? However, one problem of purely data-driven approaches is the sheer abundance of data. Even with efficient high-performance computers, it will (presumably for a long time) not be possible to process and analyse all possible configurations of a biological system. This means that we still need to guide these algorithms by formulating the right questions, implying that we still need to investigate the underlying design principles. To do so, one needs to study interaction rules and regulatory mechanisms that cause dynamic, emergent properties and functions. In terms of Aristotle’s characterisation of complexity (“the whole is more than the sum of its parts”), we need to focus on the ‘more’. To this aim, the theory of dynamic systems, including its

mathematical framework, is extremely helpful to conceptualise and to describe formally dynamic phenomena. As described above, these techniques require a simplified description of the systems. Whereas these simplifications disregard various details, they nevertheless facilitate the discovery of regulatory principles and mechanisms.

From this, it becomes clear that a systemic, mechanistic understanding of biological systems ultimately requires the combination of a detailed description of the components of the system and of the regulatory principles that rule system dynamics. Thus, data-driven bioinformatics approaches and hypothesis-driven mechanistic modelling approaches should not be seen as competing but as complementary computational disciplines.

However, any kind of mathematical model, regardless of its complexity or formalism, is based on a set of assumptions. Thus, conclusions drawn from a particular model are constrained by these assumptions, and any interpretations neglecting them can, therefore, be misleading. Importantly, this statement is not restricted to mathematical models, but holds true for experimental models, too. It is therefore important to consider the limitations and pitfalls (which are often inherent to the applied modelling strategy) to correctly interpret the results obtained from both experimental and modelling experiments.

Gene expression is often modelled as a binary (i.e. on/off) process, using Boolean network approaches. These models are extremely useful to capture the essence of the regulatory relationships, but they might also miss important details, such as the fact that relationships between TFs and target genes are dose-dependent rather than binary. Closely related to this issue, when interpreting the relationships between network components, one should also consider the properties of the particular measure used to describe the links. For example, the widely used Pearson’s correlation coefficient is constructed to detect linear, but not (in general) non-linear relationships. Thus, GRNs with links that have been deduced using this metric might potentially disregard any kind of non-linear interactions.

Another problem, which might be related to the process of adapting a computational model to experimental data, is the so-called over-fitting. It occurs if a model (i.e. its structure and/or its parameters) is adapted to one particular dataset. Indeed, a ‘perfect’ model fit to a particular set of data can always be achieved by a sufficiently large number of model parameters. Thus, it must not be interpreted as the only explanatory model but as one consistent explanation. The reliability of a model increases with the number of independent datasets or experimental conditions that it consistently explains. As different models might potentially be consistent with a certain dataset, it is suggested, according to the parsimony principle [also denoted as ‘Occam’s razor’, after William of Ockham (ca. 1287-1347): “pluralitas non est ponenda sine necessitate” (“pluralism must not be presumed except if it is necessary”)], to prefer the simplest one with the fewest assumptions. By contrast, identifying a condition that cannot be described by a given model will disprove the model in its current formulation. This highlights a very important requirement for any model: it has to provide experimentally testable predictions, because model validation strictly depends on the comparison with data.

These remarks do not imply that computational approaches are less reliable than experimental studies. On the contrary, we would like to emphasise explicitly their strengths and potentials. However, to take advantage of their application in an optimal way and to avoid potential misinterpretations, any modelling work should be accompanied by a detailed description and discussion of the

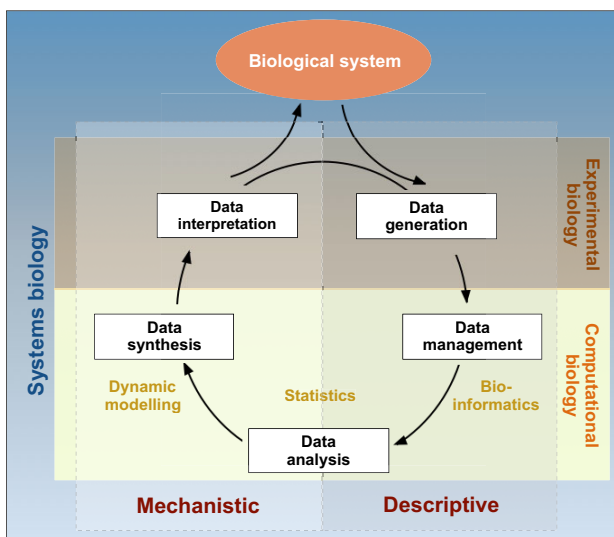


Fig. 6. Circle of knowledge generation. Schematic illustrating the general (iterative) circle of knowledge generation in biology, specifically indicating type, role and methodology class of computational approaches being addressed.

assumptions that went into the analysis. We would also like to highlight the necessity of independent confirmations of model-derived hypotheses and predictions.

Competing interests

The authors declare no competing or financial interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- Artus, J. and Hadjantonakis, A. K. (2012). Troika of the mouse blastocyst: lineage segregation and stem cells. *Curr. Stem Cell Res. Ther.* **7**, 78-91.
- Artyomov, M. N., Meissner, A. and Chakraborty, A. K. (2010). A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput. Biol.* **6**, e1000785.
- Beck, F., Erler, T., Russell, A. and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev. Dyn.* **204**, 219-227.
- Bessonard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G. and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**, 3637-3648.
- Bewick, V., Cheek, L. and Ball, J. (2003). Statistics review 7: correlation and regression. *Crit. Care* **7**, 451-459.
- Bornholdt, S. (2008). Boolean network models of cellular regulation: prospects and limitations. *J. R. Soc. Interface* **5** Suppl. 1, S85-S94.
- Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., van Oudenaarden, A. and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209-1222.
- Carter, M. G., Stagg, C. A., Falco, G., Yoshikawa, T., Basse, U. C., Aiba, K., Sharova, L. V., Shaik, N. and Ko, M. S. H. (2008). An in situ hybridization-based screen for heterogeneously expressed genes in mouse ES cells. *Gene Expr. Patterns* **8**, 181-198.
- Chambers, I. and Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. *Development* **136**, 2311-2322.
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655.
- Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-1234.
- Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* **10**, 615-624.
- Chew, J.-L., Loh, Y.-H., Zhang, W., Chen, X., Tam, W.-L., Yeap, L.-S., Li, P., Ang, Y.-S., Lim, B., Robson, P. et al. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell. Biol.* **25**, 6031-6046.
- Chickarmane, V. and Peterson, C. (2008). A computational model for understanding stem cell, trophoctoderm and endoderm lineage determination. *PLoS ONE* **3**, e3478.
- Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M. and Peterson, C. (2006). Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.* **2**, e123.
- Chickarmane, V., Olariu, V. and Peterson, C. (2012). Probing the role of stochasticity in a model of the embryonic stem cell - heterogeneous gene expression and reprogramming efficiency. *BMC Syst. Biol.* **6**, 98.
- Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M. M., Heninger, A.-K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H. et al. (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4**, 403-415.
- Dowell, K. G., Simons, A. K., Wang, Z. Z., Yun, K. and Hibbs, M. A. (2013). Cell-type-specific predictive network yields novel insights into mouse embryonic stem cell self-renewal and cell fate. *PLoS ONE* **8**, e56810.
- Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S. and Smith, A. G. (2014). Defining an essential transcription factor program for naive pluripotency. *Science* **344**, 1156-1160.
- Edgar, R., Domrachev, M. and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210.
- Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156.
- Frades, I. and Matthiesen, R. (2010). Overview on techniques in cluster analysis. *Methods Mol. Biol.* **593**, 81-107.
- Frankenberg, S., Gerbe, F., Bessonard, S., Belville, C., Pouchin, P., Bardot, O. and Chazaud, C. (2011). Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev. Cell* **21**, 1005-1013.
- Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki, J.-i. and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* **16**, 784-789.
- Glauche, I., Herberg, M. and Roeder, I. (2010). Nanog variability and pluripotency regulation of embryonic stem cells—insights from a mathematical model analysis. *PLoS ONE* **5**, e11238.
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R. and Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189.
- Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C. J., Creighton, M. P., van Oudenaarden, A. and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595-601.
- Herberg, M., Kalkan, T., Glauche, I., Smith, A. and Roeder, I. (2014). A model-based analysis of culture-dependent phenotypes of mESCs. *PLoS ONE* **9**, e92496.
- Herberg, M., Zerjatke, T., de Back, W., Glauche, I. and Roeder, I. (2015). Image-based quantification and mathematical modeling of spatial heterogeneity in ESC colonies. *Cytometry A* **87**, 481-490.
- Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853-3862.
- Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y. and Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-538.
- Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J. and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7**, e1000149.
- Karwacki-Neisius, V., Göke, J., Osorno, R., Halbritter, F., Ng, J. H., Weiße, A. Y., Wong, F. C. K., Gagliardi, A., Mullin, N. P., Festuccia, N. et al. (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell* **12**, 531-545.
- Kim, J., Chu, J., Shen, X., Wang, J. and Orkin, S. H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049-1061.
- Krupinski, P., Chickarmane, V. and Peterson, C. (2011). Simulating the mammalian blastocyst—molecular and mechanical interactions pattern the embryo. *PLoS Comput. Biol.* **7**, e1001128.
- Kunath, T., Saba-Ei-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S. and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902.
- Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431-440.
- MacArthur, B. D., Sevilla, A., Lenz, M., Müller, F.-J., Schuldt, B. M., Schuppert, A. A., Ridden, S. J., Stumpf, P. S., Fidalgo, M., Ma'ayan, A. et al. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nat. Cell Biol.* **14**, 1139-1147.
- Mainzer, K. (2014). *Die Berechnung der Welt. Von der Weltformel zu Big Data.* München: C.H. Beck Verlag.
- Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A. et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604.
- Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA* **78**, 7634-7638.
- Martinez Arias, A. and Brickman, J. M. (2011). Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Curr. Opin. Cell Biol.* **23**, 650-656.
- Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M. G., Sharov, A. A. and Ko, M. S. H. (2006). Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS ONE* **1**, e26.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642.
- Miyazaki, Y. and Torres-Padilla, M.-E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* **483**, 470-473.
- Muñoz-Descalzo, S., Rue, P., Faunes, F., Hayward, P., Jakt, L. M., Balayo, T., Garcia-Ojalvo, J. and Martinez Arias, A. (2013). A competitive protein interaction network buffers Oct4-mediated differentiation to promote pluripotency in embryonic stem cells. *Mol. Syst. Biol.* **9**, 694.
- Nakaoka, S. (2014). Multiscale mathematical modeling and simulation of cellular dynamical process. *Methods Mol. Biol.* **1195**, 269-283.

- Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N. P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M. et al. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. *EMBO J.* **31**, 4547-4562.
- Nichols, J. and Smith, A. (2012). Pluripotency in the embryo and in culture. *Cold Spring Harb. Perspect. Biol.* **4**, a008128.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391.
- Niwa, H. (2007). How is pluripotency determined and maintained? *Development* **134**, 635-646.
- Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048-2060.
- Niwa, H., Miyazaki, J.-i. and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372-376.
- Noble, W. S. (2006). What is a support vector machine? *Nat. Biotechnol.* **24**, 1565-1567.
- Pe'er, D. (2005). Bayesian network analysis of signaling networks: a primer. *Sci. STKE* **2005**, pl4.
- Peerani, R., Onishi, K., Mahdavi, A., Kumacheva, E. and Zandstra, P. W. (2009). Manipulation of signaling thresholds in "engineered stem cell niches" identifies design criteria for pluripotent stem cell screens. *PLoS ONE* **4**, e6438.
- Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. and Hadjantonakis, A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081-3091.
- Ringnér, M. (2008). What is principal component analysis? *Nat. Biotechnol.* **26**, 303-304.
- Rodda, D. J., Chew, J.-L., Lim, L.-H., Loh, Y.-H., Wang, B., Ng, H.-H. and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731-24737.
- Rossant, J. and Tam, P. P. L. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701-713.
- Schulz, H., Kolde, R., Adler, P., Aksoy, I., Anastassiadis, K., Bader, M., Billon, N., Boeuf, H., Bourillot, P.-Y., Buchholz, F. et al. (2009). The FunGenES database: a genomics resource for mouse embryonic stem cell differentiation. *PLoS ONE* **4**, e6804.
- Silva, J. and Smith, A. (2008). Capturing pluripotency. *Cell* **132**, 532-536.
- Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737.
- Singer, Z. S., Yong, J., Tischler, J., Hackett, J. A., Altinok, A., Surani, M. A., Cai, L. and Elowitz, M. B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319-331.
- Singh, A. M., Hamazaki, T., Hankowski, K. E. and Terada, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25**, 2534-2542.
- Smith, A. G. (2001). Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol.* **17**, 435-462.
- Smith, A. (2013). Nanog heterogeneity: tilting at windmills? *Cell Stem Cell* **13**, 6-7.
- Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093-2102.
- Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676.
- Takahashi, K. and Yamanaka, S. (2013). Induced pluripotent stem cells in medicine and biology. *Development* **140**, 2457-2461.
- Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Sonogr.* **6**, 35-39.
- Torres-Padilla, M.-E. and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development* **141**, 2173-2181.
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909-918.
- Trott, J., Hayashi, K., Surani, A., Babu, M. M. and Martinez-Arias, A. (2012). Dissecting ensemble networks in ES cell populations reveals micro-heterogeneity underlying pluripotency. *Mol. Biosyst.* **8**, 744-752.
- van den Brink, S. C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.-K., Nowotschin, S., Turner, D. A. and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231-4242.
- Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W. and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-368.
- Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. and Brivanlou, A. H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847-854.
- Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. *FASEB J.* **11**, 835-841.
- White, D. E., Kinney, M. A., McDevitt, T. C. and Kemp, M. L. (2013). Spatial pattern dynamics of 3D stem cell loss of pluripotency via rules-based computational modeling. *PLoS Comput. Biol.* **9**, e1002952.
- Wray, J., Kalkan, T. and Smith, A. G. (2010). The ground state of pluripotency. *Biochem. Soc. Trans.* **38**, 1027-1032.
- Xu, H., Barouk, C., Dannenfels, R., Chen, E. Y., Tan, C. M., Kou, Y., Kim, Y. E., Lemischka, I. R. and Ma'ayan, A. (2013). ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. *Database* **2013**, bat045.
- Xu, H., Ang, Y.-S., Sevilla, A., Lemischka, I. R. and Ma'ayan, A. (2014). Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput. Biol.* **10**, e1003777.
- Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49-52.
- Yamanaka, Y., Lanner, F. and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715-724.
- Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523.