Transition states and fate decisions at branching points in Waddington landscapes

Naomi Moris[1], Cristina Pina[2] and Alfonso Martinez Arias[1*]

1. Department of Genetics
University of Cambridge
Cambridge, CB2 3EH
UK
2. Department of Haematology
University of Cambridge
Cambridge, CB2 0PT
UK

* Author for correspondence
email: ama11@hermes.cam.ac.uk

Preface

The Waddington landscape is an abstract metaphor frequently used to represent the relationship between gene activity and cell fates during development. Over the last few years it has become a useful framework for interpreting results from single-cell transcriptomics experiments and has led to the proposal that during fate transitions cells experience smooth, continuous progressions of global transcriptional activity, captured by (pseudo)temporal dynamics. Here, focusing strictly on the fate decision events, we suggest an alternative view: that fate transitions occur in a discontinuous, stochastic manner where signals modulate the probability of the transition events.

Introduction

The development of a multicellular organism with its array of organs and tissues is a reproducible event that, upon detailed observation, reveals a coordination between two processes: an increase in cellular mass and phenotypic diversification of the expanding cell populations. Classical studies of organisms with fixed lineages and numbers of cells, established that there is an order to this process: cells adopt different identities with exquisite spatiotemporal precision in parallel to the increase in numbers[1,2]. In the case of *C. elegans*, the degree of this association is extreme, as the outcome of every cell division is largely predictable in terms of the identity, fate and relative position of the emerging cells[1]. This reproducibility suggests the existence of an underlying programme, a sequence of instructions towards an end point which in this case are specific cell types or fates. These programmes are implemented by **Gene Regulatory Networks** (GRNs); fundamental units of molecular activity which build and maintain functioning tissues by promoting sequential, and largely irreversible patterns of gene expression that link genes to cellular lineages[3,4]. The universality of this notion is supported by the observation that tissues and organs of insects and vertebrates, which do not have fixed numbers of cells, also emerge through lineages, in these cases polyclonal[5], associated with ordered sequences of gene expression[3,6,7].

Programmes of gene expression are cell-autonomous, in that they operate independently and reproducibly in individual cells. An example of this is the lineages of stem cells in several tissues[8,9] and, most clearly, in the differentiation processes of Embryonic Stem (ES) cells which follow closely the events in embryos[10,11]. A formal consideration of the relationship between cell lineages and genetic programmes during development highlights some essential elements of any process of **cell fate** specification: 1) a sequence of instructions associated with a given fate, 2) specific decision events, 3) directionality and 4) a means of apportioning defined numbers of cells to particular fates to generate proportionate tissues. The most crucial feature of any cell fate specification process is the point at which cells make choices, the so-called *cell fate decision*. For the most part, this process is binary[12] and steers cells to adopt one of two states, thus enabling an isogenic population to assume divergent states over time, each with a specific profile of gene expression.

Many attempts have been made to understand the connection between programmes, fates and decisions in developing organisms and to identify the underlying molecular mechanisms. One of the most enduring is Waddington's notion of the Epigenetic Landscape[13,14], which focused on the activity of single cells and tried to conceptualise the emergence of developmental choices as the result of intrinsic constraints (regulatory interactions) shaped during evolution. Waddington drew a number of representations of his famed landscape[15] but two (Figure 1A and 1B) have come to be identified with his views on development and evolution. In the first and most popular one, a cell, reimaged as a pebble, begins at the top of a hill and follows existing paths in the landscape driven by a gravitational force which leads it into one of several possible fates represented as valleys. In principle, the decisions about fate are stochastic, so if there are many cells rolling down the landscape they will go one way or another independently. Once a cell makes a decision, it is restricted in its subsequent decisions by the route it has taken, representing decreased 'cellular potential' and fate restriction. In a second image, Waddington presciently implies that there are genetic regulatory mechanisms responsible for the shape of the landscape. Any modern interpretation of Waddington's landscape needs to consider three elements that are built into his vision: the discreteness of the final states or fates, the smooth paths down which cells travel as they move towards these states, and the decision events which occur at binary branches in the landscape.

Although Waddington was aware of some **dynamical systems** theory, he used the epigenetic landscape more as a metaphor for development, without being explicit about its details. A formal underpinning of the landscape was first suggested by S. Kaufmann in 1969 who formalized the landscape by identifying the valleys with *attractors* in Boolean networks[16,17]. Later, S. Huang developed this notion further into a multidimensional *dynamical systems* framework[18,19]. In this formalism, Waddington's intuition of genetic control of the landscape is brought to light using gene expression profiles projected onto an n-dimensional phase-space with vector fields where stable states act as attractors[16,18]. It is these frameworks that have perhaps become most popular when invoking the Waddington landscape as a reference to interpret gene expression data in development and disease.

There is no doubt that the Epigenetic Landscape is a compelling and prescient analogy for how the static information encoded in the genome is translated dynamically into tissues and organs and how it is used to drive cell fate decisions. However, despite its popularity, when examined in closer detail, it harbours problems of formalism and interpretation. While a good metaphor, it is difficult to implement formally; **potentials** cannot realistically be calculated analytically for more than 1 or 2 dimensions[20], so interpretations of the z coordinate in graphical outputs need to be treated with care, especially since cellular states are often characterised by differences in many thousands of genes. More specifically, as a system far from chemical equilibrium, with no conservation of energy (non-equilibrium), and where the energy effort to move from A to B depends on the exact path, rather than the difference in elevation (non-gradient), the true potential landscape is unobtainable and alternatives (discussed later) require some assumptions. Furthermore, an essential feature of developmental processes is their dynamic behaviour with properties that can make their accurate representation difficult to implement. For instance, developmental systems often show overall directionality with some local reversibility, associated with high levels of control such that cells can make decisions that are seemingly deterministic rather than stochastic and with specific temporal dynamics. But the most critical challenge to Waddington's vision arises specifically at the branching points in the landscape, where cell fate decisions take place. In Waddington's representation, although the final states are clearly discrete, at decision points, smooth channels simply branch-off as states diverge, implying that cells passively 'slide' into the decision event (Figure 1A & 1B). Alternative theories posit that the initial and final states do not converge and that the decision is discontinuous[21] or that cells must actively 'jump' to overcome barriers between discrete fates[22]. This distinction has broad implications for the interpretation of experimental datasets at cell fate decisions. With the rise in single-cell transcriptional methods and the accompanying statistical and computational analytical tools, we have more ability than ever before to begin to unravel some of the complexity of cell fate decisions and to challenge longstanding paradigms.

Here, we consider these issues and discuss the relationship between Waddington's Epigenetic Landscape and genetic programmes in the light of recent advances in genomics and, specifically, single cell technologies. We focus particularly on the decision events within cell lineages, reviewing the relationship between the data and the abstract representations that exist for its interpretation, using examples and illustrations from recent literature. We suggest that there might be a fundamental flaw in the reinterpretation of Waddington's landscape in terms of an assumption of continuous trajectories that branches smoothly at cell fate decisions, and instead propose the notion of a 'transition state' as an alternative which accounts for the gene expression **heterogeneity** in terms of discontinuities in the mechanisms of fate choice.

Gene Regulatory Networks and Cell State

The genetic analysis of development in model organisms led to the identification of genes associated with particular cell fates and, through epistasis analysis, to the arrangement of those genes into 'developmental pathways' such as the segmentation cascade of *Drosophila*[23,24] or the vulva pathway in *C. elegans*[25]*.* The interpretation of these pathways rests on two important assumptions which permeate gene expression analyses of similar processes: first, since development proceeds over time, the chains of gene activity must reflect continuous temporal sequences of events and secondly, that the process of diversification proceeds through binary choices. As different studies uncovered an increasing number of genes involved in particular processes and identified many of them with regulatory transcription factors, pathways became associated with the activity of GRNs[3]. In its basic structure, a GRN captures a collection of interactions between transcription factors and their targets, and can be represented in network diagrams with genes or proteins as nodes, with their input/output relationships as edges which, when possible, have a directional component[26]. The signs (+,-) and strengths of the edges are identified either genetically or biochemically and provide a rational description of the regulatory relationships of a system. In its ideal form, a GRN should contain a complete molecular description of a specific process and provide insights into the manner in which multiple transcription factors create a unique 'code' for a particular cell fate and, if possible, for cell fate decisions. However a complete GRN - as the sum total of transcription-level factors associated with a fate - is impractical to model; just deriving all the parameters of each interaction across a transcriptome is an insurmountable challenge in most biological systems. Fortunately there are ways to reduce this complexity and, in the process, potentially identify the significant elements that drive biological events. The ability to do so lies in the observation that GRNs are organised in hierarchical and highly modular structures[27-29].

Typically, high-throughput transcriptional studies generate large correlational structures, which although useful to identify markers or signatures of various cell types, are difficult to link to dynamic processes. Alternatively, such data can be used together with perturbation experiments to identify small, regulatory circuits of the system, often arranged in common structural patterns known as network motifs[26,30] which are far more amenable to modelling. The simplest GRN structure which might give rise to a directional state transition event is linear regulation, where a signalling input increases Gene A, which regulates Gene B expression and so on, meaning that an input regulates a cascade of downstream genes resulting in the equivalent of a 'gene avalanche'[20]. While this model is sufficient to cause systemic directionality, it proves lacking for other contexts where signalling, rather than having a singular defined cell type output, works pleiotropically to regulate diverse fate decisions at varying times during development. Instead of simplistic linear structures, there is evidence for core feedback loops in regulatory networks controlling development, effectively creating circular structures that are adapted to produce stable, discrete cellular states[29,31,32]. Thus, GRNs can be constructed as interactomes of network motifs, which provides a more insightful and functional representation of the processes under consideration (see reference[30]).

Waddington's Potential Landscape

An advantage of a network motif formalism of GRNs is that landscapes similar to Waddington's Epigenetic Landscape can be largely reconstructed using dynamical systems representations (Figure 1). For example, bistable switches have many of the properties required for stable fate decisions and have been implicated in various systems where a decision results in two discrete states[33-35]. Such network motifs can be modelled and projected onto a phase-plane that identifies the stable points of the system and the vectors that govern the motion of any state. If one could calculate the potential of every position on the phase-plane, an extra dimension could be added that would generate a landscape. However, it is important to remember that for non-equilibrium, non-gradient systems such as genetic networks, where the dimensions are greater than one, it is not possible to obtain a strict potential function[20]. The alternative is to evaluate the steady-state probabilities of each state or to decompose the vector field and approximate the *quasi*-potential landscape[19,20]. Using these methods, a value of *quasipotential* can be calculated for every state which is inversely correlated to the probability that a cell will assume that state, meaning local 'low-energy states' are more densely populated[36]. Changes in the expression profile of any cell are then associated with altered coordinate positions meaning dynamic state transitions can be directly mapped onto the landscape.

In the extreme, there are two means by which cells can change their state: by stochastic fluctuations that cause cells to jump around the landscape without requiring any parameter changes, or by extrinsic inputs to the system that alter parameters of the network and change the landscape geometry (the position, shape or size of landscape elements) or topology (the *number* of landscape elements; Figure 1D)[37,38].

Stochastic fluctuations appear to be an intrinsic feature of some biological systems, especially with regard to transcription where low numbers of transcription factor and DNA molecules can cause sporadic engagement of RNA polymerase and lead to discontinuous mRNA synthesis[39,40] resulting in deviations from the stable attractor state. Usually these perturbations are small enough that the force of the attractor will draw the system back to its original stable state, but occasionally cells might cross a barrier in the landscape that divides attractor basins, causing a spontaneous state-switching event. Biologically these events could represent stochastic fluctuations, perhaps of transcription factor levels, over a threshold that proves sufficient to promote cell identity changes[41]. Such events have been described in bacteria where they have been shown to have functional significance[42]. Stochastic fluctuations might also apply to stem cells in eukaryotes and, in particular, embryonic stem cell populations[43-45]. Such noise-induced transition events can be modulated by the landscape geometry such that deep attractor basins, or high barriers between basins, can reduce the probability of spontaneous transition events and even impose directionality on the transitions[20,46].

Alternatively, the change in state can be associated with a transformation of the landscape, associated with variations in some parameters of the system[38] and often mediated by input stimuli on the networks. In a biological context, these inputs could represent extracellular signalling associated with cell fate decisions like *Fgf, Nodal* or *BMPs.* At critical parameter values, the landscape can change qualitatively, for instance converting a monostable system, with one stable state or attractor, to a bistable system, with two. These dynamical systems changes can be related to bifurcations (see Box 1). This formalism is important as it can predict systemic behaviour from an underlying GRN, providing a testable assessment of various models. Identifying which model best approximates the biological behaviour of cell fate decisions requires experimental evidence of sufficient resolution as to be able to test predictions, ideally utilising dynamic data from single cells.

Single-cell Resolution of Fate Decisions

Over the last ten years there have been remarkable advances in our ability to monitor the transcriptional activity of single cells either in multiplexed qRT-PCR for specific sets of genes or in whole transcriptome RNA-seq [47,48]. The output of these experiments is a gene expression matrix of *n* genes across *m* cells revealing how the genome is expressed within individual cells, from which the corresponding distribution pattern of expression across a population can be estimated. A complementary technique measures protein levels at single cell resolution by labelling antibodies with heavy-metal tags (CyTOF)[49]. Applications of these technologies to an ever increasing number of case studies have revealed that although a cell type is indeed largely characterized by differential expression of particular genes, when observed at the level of individual cells even phenotypically homogeneous cell types display a high degree of heterogeneity in the expression of individual genes. Indeed, the crucial difference between ensemble and single-cell level transcriptional assays lies resolutely in the ability to observe heterogeneity; the challenge is to interpret this heterogeneity in a meaningful way (reviewed in references[48,50,51]).

Sources of observed heterogeneity can be broadly divided into two groups; those associated with the experimental protocol or those that are an integral element of the process under study. While technical noise is indeed a concern in the analysis of the data, accurately quantifying technical noise remains challenging. Yet there are experimental and analytical means of reducing the impact of such occurrences including RNA spike-ins of known concentrations, or through application of compensatory models[52,53]. If suitably accounted for, the remaining variability can be associated with biological processes which may or may not have functional relevance. Over the last few years, mouse ES cells have been well-studied in this context, and there is now ample evidence that, in this system, broad distributions of many transcription factors have functional significance[44]. In the case of Nanog, for example, its expression distribution has been shown to be dynamic, with individual cells capable of reforming the original distribution, suggesting some overall control of the heterogeneity[43,54,55]. Importantly, cells with different expression levels of Nanog have different functional capabilities for self-renewal and differentiation[43,56], implying that dynamic heterogeneity is functionally relevant to cellular decision making.

Attempts to ascribe meaning to observed heterogeneities in gene expression, alongside the shear growth in popularity of single-cell methods, have necessitated the development of statistical and computational analyses capable of accessing the wealth of information harboured within the patterns of these heterogeneities. Many of these techniques were borrowed from other disciplines while others have been devised specifically for the application to such datatypes (Box 2). The simplest methods seek to cluster cells based on similarities of their expression profiles and distinguish them based on differences, thereby revealing underlying structures within the population. Many of these methods execute weighted dimensionality reductions that turn the n-dimensional space into a number of components or dimensions that are more readily visualised, by identifying and retaining the information on the population structure. Classification of single cells into clusters by mapping them to gene expression patterns can identify subpopulation structure within the dataset. Clustering analysis of single-cell data is particularly valuable in identifying very rare or transient populations of cells that might not have been previously observable from low-dimensional or ensemble assays[57,58].

Subpopulation characterisation based on multi-dimensional gene expression data can also relate back to original estimates of population structure based on alternative methods, to validate or challenge it. In haematopoietic progenitors, sorting strategies based on cell surface markers are often used to capture distinct functional cell types[59,60]. However, a recent study identified new myeloid progenitor subpopulations from single-cell gene expression clusters and suggested that traditional sorting strategies might result in mixed populations, and showing that the transcriptionally-defined subpopulations better predicted the functional capacity of cells[61,62]. These findings provide a transcriptional regulatory framework in which to place observations made a decade ago that challenged the routes of lineage specification[63-65] suggested by the classical haematopoietic hierarchy[66].

But the power of single-cell transcriptional data does not only lie in its potential to reveal structures within populations, but in the possibility that it contains information about the *dynamics* of the gene regulatory networks during fate decisions. Assuming that gene expression profiles provide an estimate of the coordinates for a phase space of the dynamical system which defines the state of a cell, heterogeneities can be used to infer the mechanisms of transitions between different states of a system. The potential of this approach has been shown using a stochastic model of lineage commitment in the hematopoietic system[67]. By using the combined transcriptional patterns of three transcriptional regulators in individual self-renewing and committed cells, the model calculates the probability of transition from self-renewal to the committed state associated with each individual transcriptional pattern, and successfully recapitulates *in silico* the dynamics of a differentiation culture system. In this example, cells assayed came from a source in constant asynchronous flux, such that cells were observable at all points of the dynamic process. This makes a key assumption of ergodicity; that a snapshot of cell identities at one time point is equivalent to a longitudinal observation of one cell over time. In other examples, cells originate from relatively synchronised populations with observations recorded at different temporal intervals, such as different developmental stages or times after a differentiation cue.

When time resolved data has been collected, the major challenge is to extract the salient features in a manner that reveals the underlying dynamic processes (Figure 2). These challenges are ongoing, but progress has been made towards this end with recently devised algorithms derived from dimensionality reduction techniques that have been used to create a sequence of ordered events in a dimension referred to as '**pseudotime**' (two of the most common are Wanderlust[68] and Monocle[69], see also[70]). These methods seek to minimise the variability conferred by cellular heterogeneity through ordering cells by similarity (Figure 2B). As such, any heterogeneity of cells during a decision is minimised as average continuous trajectories are drawn. Any remaining observed heterogeneity between cells at a snapshot in time might then be reduced to asynchronous traversal of a fixed pathway or lack of synchrony in decisions. These methods have already been applied to systems including primary human neuroblasts[69], human B cell lymphopoiesis[68] and haematopoietic stem cells[71].

It is unclear how much the output of these methods tells us about the specific process of cell fate decisions, since there is a considerable assumption which influences the pseudotime interpretation: that the transition between states follows a continuous trajectory. This view has been compared to a temporal interpretation of Waddington's landscape, but it is not implicit in the original formation of the landscape and need not reflect the actual course of events, particularly at decision events where changes of state might be mediated by discontinuous mechanisms which would be obscured by methods attempting to linearize heterogeneities into a continuous, convergent pathway.

The Transition State

Pioneering single cell studies of fate decisions in the hematopoietic system revealed that cells with multilineage potential could co-express genes typically associated with each of their alternative lineage fates[72]. For the most part, expression was infrequent and at low levels that varied from cell to cell[72]. Similar observations have since been made in several different systems[73-76], specifically in single cells within populations undergoing fate changes[77,78] and, notably, in stem cell populations where they are thought to play a role in the balance between differentiation and self-renewal[79-81]. These observations suggest that heterogeneities might represent a general feature of changes of state[44,82,83] and has led to the proposal that heterogeneous patterns of expression are associated with the cell fate decision event at the level of single cells[44,72,84-86]. As such, the varied expression of genes associated with each alternative fate in single cells can act as a substrate for selection by signals[87], or as an exploration of a phase space where changes in the levels of the regulators can lead to stochastic fate change[67,81]. In either case, when cells make a decision, they upregulate the expression levels of the gene cohort of their chosen fate and downregulate those of the alternative one, as has been seen in many systems[88,89].

In the context of the premise that fate decisions typically occur between discrete states or 'attractors', the heterogeneities in gene expression as cells change state suggest a general principle*:* cells within a discrete attractor state may experience a degree of transcriptional stochasticity that can, with different probabilities, result in various transcriptional profiles characterised by the expression of genes associated with one or more independent cell identities. Such expression profiles which are transient, distinct from cell to cell and manifest as heterogeneities at the population level, endow cells with varied probabilities of effecting a cell fate transition. We have called this collection of transcriptional profiles a '**transition state**' (TS), and suggest that it represents a substrate for cell fate decisions by facilitating state switching whilst retaining the reverse transition probability (Figure 3). An important element in the notion of the transition state is that the passage from one state to another need not be smooth and continuous as portrayed in representations of pseudotime. Furthermore, a transition state can be characterized by bi- or, in certain instances, tri-stability since a cell can have access to either two or, if they still express some genes of the initial state, three states.

The term 'transition state' refers to an analogy with the well-known *transition state* in chemistry, introduced to provide a mechanistic underpinning for the progress of a chemical reaction[90]. The induction of a reaction between two substances, by a new reactant or a catalyst, triggers a short-lived intermediate in which the molecules involved engage into a number of configurations exhibiting intermediate characteristics between reactants and products. It represents a state of maximal potential energy as a chemical potential barrier is traversed, often using catalytic factors to reduce the intervening barrier height (Figure 3A). In an analogous fashion we surmise that during a fate transition, a GRN can be seen as a reactant that receives a new input, from a signal or from crossing a threshold of its own activity, that leads to either a new network or to a new pattern of connectivity of the existing network. By analogy with a chemical reaction, the cell fate decision process has an intermediate during which the networks explore the state space (Figure 3B) and this manifests as heterogeneities in gene expression at the level of single cells. While the analogy between the chemical and biological transition states holds value in its description of an intermediate state, care should be taken not to extend the analogy too far, especially since the biological transition state is limited by its thermodynamic properties that prevent a formal calculation of potential.

A corollary of the notion of the transition state is that any given phenotypic state might not be associated with just one network state. If GRNs involved in cell fate decisions are hierarchical, it might be that there are key configurations of specific network motifs or small networks that are sufficient to trigger a particular state, independently of additional or downstream gene activity. Pursuing further a chemical analogy, a cell state or fate can be described by its '**macrostate**' (the core GRN connectivity that gives rise to a specific phenotypic state) which can be represented by more than one '**microstate**' (small changes to the expression profile, which do not impact on the phenotypic state; Figure 2C)[91]. The number of potential microstates depends on the number of nodes of the network, but whether a cell will assume any given microstate is dependent on the network configuration and its biological relevance (for instance, a cell can only assume a positive value of a transcription factor level). In a transition state, cells may explore a higher number of microstates. This notion has been formally discussed in the context of mouse ES cells but is, in principle, applicable to any cell fate transition.

Landscapes and transition states

The notion of the transition state leads to the corollary that the heterogeneities in gene expression, observed with single-cell transcriptomics in systems undergoing fate decisions, need not align through time along an ordered continuum, and could instead be interpreted as a reflection of systemic features that allow a cell to stochastically rearrange its networks by exploring the local phase space. We do not venture to suggest that cellular trajectories do not exist at all; merely that at the decision event, the actual transitions between states might be discontinuous, and that the observed transcriptional heterogeneity reflects the existence of a dynamic array of transcriptional states, with varied probability of transitions. It is highly likely that downstream of the decision, external signalling or intrinsic programmes of the GRN result in continuous trajectories through geometric landscape changes leading to the next decision point in a manner that is consistent with the temporal continuity implicit in the notion of pseudotime. The distinction between the two notions arises at the point where cellular decisions are made (Figure 2). There are hints that discontinuous fate decisions do occur in some contexts, for instance in the hematopoietic system, particularly in the case of early myelo-erthroid choices in mouse haematopoiesis[81] and, in human haematopoietic stem cells that can give rise to differentiated cells directly[92] without going through multi-lineage intermediates. But the framework could be applied to many binary decisions, such as the neuromesodermal progenitor in the mammalian embryo which generates neural and mesodermal progenitors from a common precursor which coexpresses neural and mesodermal genes[93,88].

The notion of a 'state' , the transition state, made up of a collection of cells, each with a different transcriptional profile that are disconnected temporally is, at first sight, at odds with the intuitive, smooth and continuous interpretation derived from a cursory examination of Waddington's landscapes. However, by representing the process in terms of a computable dynamical system with parameter changes, we can apply **bifurcation theory** to examine systemic behaviours[37,38,94], particularly focusing on bifurcation events (see Box 1). In this framework, a Subcritical Pitchfork Bifurcation (see Box 1 and Huang, et al. [38]) captures many of the features of cell fate decisions that we have described as a transition state. Namely, the specific range of parameter values conferring tristability could correspond to a transient 'window' where transitions are promoted; the system assumes a permissive topology in which both the initial and final states stably coexist and transcriptional exploration is favoured since the landscape assumes relatively high energy (quasipotential) stable states with low intervening barriers. This bears a clear resemblance to our proposed transition state, as the landscape promotes increased heterogeneity across the population due to the relatively low quasipotential force of each attractor. Like the Supercritical Pitchfork Bifurcation, the initial and final states are discrete, suggesting transitions might occur through state-switching events rather than continuous divergence of states as implicated in Waddington's Epigenetic Landscape.

Importantly, the bifurcation diagram as drawn in Figure 1D does not represent the evolution of the system as a function of time, but as a function of a parameter or parameters which could themselves also vary with time, not necessarily in a deterministic or linear fashion[37]. An example of these parameters could be extracellular signals. This means that the same system could easily be tuned to different circumstances, for instance by applying high cooperativity to the input that alters the system's parameter which would effectively reduce the time the system spends in the tristable regime. The inverse is also possible, with certain systems potentially tuned to elongate this transition-promoting phase, a possible example of which are embryonic stem cells. These cells exist in a pluripotent state that, in the developing embryo represents the transition to differentiation, is very short lived and not renewed, but which is stabilized and maintained indefinitely in certain culture conditions[95,96]. These cells might then be described essentially as a 'trapped' transition state[12,85,97].

A molecular interpretation of the Transition State

At the molecular level, a cell fate decision necessarily involves a change in the transcriptional state of a cell. Gene transcription requires a close interaction between three sets of components, each of which involves large multiprotein complexes: (1) a suite of transcription factors that define a state by promoting spatiotemporal control of gene expression; (2) histone modifiers and chromatin remodellers, that determine the accessibility of the transcription factors to the DNA and their binding stability, and (3) the basal transcriptional machinery, which executes the transcriptional process and, to a first approximation, is unlikely to vary throughout the process. Proteins that interact with each and any of these machines have the potential to change the coordinates of the transcriptional state of a cell, but the effects, and the type of input that an individual protein will have on the process, are different in each case. For example, transcription factors can define the cellular state by activating specific genes and thus define which networks will be active; this could be a noisy process if the numbers of molecules of the participant proteins are limiting and vary from cell to cell[39,40,98-100] thus creating a substrate for regulation. The effectors of BMP/TGFß signalling, the Smad proteins, would fall into this class and their activity may result in a change in the parameters of the GRN as on their own they have weak outputs that are stabilised by association with other transcription factors[101,102], that change their binding kinetics, and by extension, the associated landscape. This is an example of how an interaction between the intrinsic cellular state (defined by extant transcription factors) and extrinsic signals (BMP/TGFß in this case) can sculpt the outline of the landscape in a synergistic manner: the interaction between the two results in the expression of a set of genes that neither of them alone can activate, and thus changes the landscape.

The role of the other components in fate decisions is less specific. They do not determine a cell state *per se*, but rather govern the robustness of the transcriptional process by modulating the binding kinetics of particular transcription factors (chromatin modifiers) or altering the frequency and amplitude of the transcriptional process (basal transcriptional machinery). Thus, any event that targets any of these will affect the efficiency of transition between states in individual cells[45]. In landscape terms, such interactions might change the potential barrier i.e. act as the catalysts in a chemical reaction, and on the basis of their strength determine the number of cells that would change state over time. We have suggested that Wnt/ß-catenin and FGF/ERK signalling pathways act in this manner[12,41,45]. This level of regulation is particularly important at the decision events and would allow tuning of the number of cells making fate decisions.

The transition state in developmental populations

A valuable feature of the proposed framework is that it provides a mechanism for the control over the size of any given population, because some of the variables can act to bias the landscape towards certain fate decisions. This is because whilst decisions are taken at the level of single cells, at the population level, biases in the bifurcation landscape will result in differences in the number of cells apportioned to each state. Extrinsic (e.g. signalling) or intrinsic (e.g. basal transcriptional machinery and chromatin modifiers) factors that bias state-space exploration towards specific transcriptional states, or that change the probability of transition associated with each individual states, can ultimately alter the final number of differentiated cells and their relative proportions. This poses a potentially useful framework for the homeostatic regulation of the size of a population[86].

An example of this situation can be found in the partitioning of the ICM of the preimplantation mouse embryo into the embryonic epiblast and the extraembryonic Primitive Endoderm. This binary decision occurs in a population of cells deemed to be in a transition state as cells express variable levels of genes associated with both fates[103-105]. The heterogeneities are resolved through the activity of a small transcriptional network involving GATA factors and Nanog, which is modulated by FGF signalling. The decision can be recapitulated in ES cells where the mechanisms involved in the decision can be explored in detail[41]. Modelling of this decision has suggested that it is driven by a tristable[106,107] network that can be reduced to a bistable one[41], which accounts for most of the observations. The outcome is two populations with balanced cell numbers that will interact later in development to steer the patterning of the embryo. How the precise and reproducible partitioning is coordinated is not fully understood, but there is evidence that it is achieved through time-integration of FGF signalling by the transcriptional network. A population in a transition state is an ideal substrate for this integration as it has an *a priori* equal probability of each cell adopting one of the two alternative fates. This system acts as a good model for similar decisions that occur during development and which result in precise and robust partitioning of populations.

It will be interesting to see how general and useful the notion of the transition state is, not only in terms of its ability to represent data but, significantly, in its ability to frame the role and activity of signals at the decision events. Whereas in most representations of developmental decisions signals simply contribute another node to the network, in our view they represent parameters that can change the structure and the dynamics of the landscape that cells explore. To test these ideas we shall have to go beyond the analysis of snapshot gene expression data and explore the dynamics of individual genes within individual cells. A validation of the hypothesis of the transition state will probably require a combination of quantitative live-imaging of cellular dynamics through a decision event, with concurrent measurement of multiple genes predicted to be indicative of transition events. In the case of a continuous transition, as portrayed by pseudotime, cells at the same point of the fate transition will be transcriptionally very similar to each other while in the case of the discontinuous one, they should be very dissimilar. By observing cells at different stages, according to the continuous pseudotime representation, all sampled intermediates between functional states will follow an identical sequence of cellular states. By contrast, the transition state theory predicts that cells will transition between states through a diverse array of dynamic mechanisms (see Figure 2).

We believe that the discussion of the contrasting frameworks that we have raised is important, since what is at stake is not simply a representation of the data derived from single cell analysis, but the mechanisms underlying fate decisions and their impact on assigning different fates to cell populations which create the building blocks for tissues and organs. In addition we hope that this discussion will open up a re-consideration of the role of signals in fate decisions and how they impinge on the dynamics of cell populations rather than just on the fates during development. Importantly, we want to encourage an open-minded view of Waddington’s epigenetic landscape which, over the last few years, has become an icon for cell fate decisions in homeostasis and development.

Glossary

Heterogeneity (*in gene expression*): variability in the expression of a gene or a group of genes across a population at single-cell resolution.

Gene Regulatory Network (GRN): a unit of interacting proteins functionally constrained by defined regulatory relationships that provide a structure and determine an output in the form of a pattern of gene expression. It is usually represented by nodes (proteins) and edges (their relationships).

Genetic Programme: temporally ordered interactions between proteins, usually transcription factors, associated with the emergence of cell types.

Cell Fate: the developmental destination of a cell if left undisturbed in its environment. The fate of a cell is more restricted than its potential.

Cell State: the transcriptional output of a gene regulatory network, with a variable degree of stability; development is characterised by sequences of cell states that culminate in specific fates.

Dynamical System: a system defined by collection of related variables that evolves in time according to certain rules. A gene regulatory network is an example of a dynamical system in which the variables are the transcription factors that represent the nodes.

Potential: ***Biologically***, the range of fates into which a cell can develop. It is reduced during development and is obscured by mere observation of events in, for example, lineage-tracing experiments, which only reveal fates. ***In physics***, potential can be described as the ability to do work and represents an amount of energy stored for that purpose. In both Biology and Physics, it represents an ability to do something.

Bifurcation theory: branch of mathematics associated with dynamical systems which accounts for the evolution of a physical or biological system according to a control parameter.

Pseudotime: notion derived from the analysis of single cell transcriptomes in cell population which allows the ordering of individual cells based on minimal differences of their transcriptomes. It has an implicit assumption that the resulting order reflects a smooth and continuous change in the state of the cell and aims to relate this change to changes in gene expression.

Transition State: an intermediate during cell fate decisions in which a cell exhibits a mixed identity between two or more states, which often represents the state of origin (i.e. the initial state the cell is in) and that of destination (i.e. the identity that the cell is adopting). It is highly unstable and reversible.

Macrostate: notion derived from statistical mechanics which defines the macroscopic state of a system e.g. a particular volume or temperature and, in the case of a biological system, a *functional* state.

Microstate: notion derived from statistical mechanics that defines a configuration of the elements that are associated with a particular macrostate of the system e.g. a molecular configuration associated with a particular volume or temperature. Any given macrostate may be associated with many different microstates. We surmise that gene expression profiles can be related to microstates in a biological context.

Box 1 | **Bifurcations and Cell Fate Decisions**

A Gene Regulatory Network (GRN) can be construed as a *dynamical system* in which the temporal evolution of the system is a function of the parameters of the network (e.g. the rate constants of the gene interactions). An important consequence is the ability to apply ‘bifurcation theory’, a geometrical formalism that identifies critical parameter values at which qualitative changes in systemic behaviour occur. This is particularly important in the context of fate decisions, since such analysis could identify the drivers of the system and dynamics of the decision making process.

There are several types of bifurcation[108], each with specific features that produce different behaviours. Here we focus on three that are pertinent to the study of cell fate decisions[37,38,94]. The simplest is the ***Supercritical Pitchfork Bifurcation*** (SrP) which maps well to the classical Waddington landscape, since it allows for a temporal diversification of cell states by binary fate decisions. Cells in a locally monostable regime on a landscape undergoing a *SrP* would gradually move into one of two new stable states after the bifurcation and as a result, fate decisions appear smooth and continuous (in the mathematical sense), since a cell remains in a stable basin throughout the decision event. While this bifurcation type is favoured by many models of fate decision, it is crucially unable to account for a key feature of cell fate decisions, namely their irreversibility[94] (see Main Text). In the *SrP*, the temporal dynamics of the landscape depends entirely on the parameter change: if the parameter were reversed, cells from both of the two final states would simply slide back to the initial state. This reversibility is a challenge in the context of cell fate decisions since state reversibility is not often observed, instead requiring experimental manipulation.

Alternatively, irreversibility can be achieved through a ***Saddle-Node Bifurcation***[94] (*SN*). In this case, rather than the *creation* of new stable states, all stable states pre-exist in the system and a state is *removed* as the parameter value changes, when an unstable point (the ‘saddle’) and a stable state (the ‘node’) meet, effectively destabilising the initial state. Similar to the *SrP*, the *SN* is capable of reproducing the temporal progression of cell states, but it also has the intrinsic properties of irreversibility and hysteresis because reversing the parameter levels in a deterministic system will retain cells in their final state, since a barrier between states exists. Importantly, the initial and final states (A and B/C) never converge, meaning the temporal state change is itself discrete.

A third type of bifurcation, the ***Subcritical Pitchfork Bifurcation***[87,108] (SbP) contains elements of both the *SrP* and the *SN* and has clear parallels with the cell fate decision process. Like the *SrP,* this bifurcation begins with a locally monostable regime (A) and resolves into a bistable regime (B and C), except that it does so through a transient period of tristability, initiated through two *SN* bifurcations. For a range of parameter values, there is therefore an overlap between the initial and final states. This model has interesting features, including directionality and hysteresis (i.e. a cell that transitions to a final state reverts back to the original state at a different parameter value to the one at which it transitioned). But the key feature of this bifurcation is that in a stochastic system, there will be heterogeneity and local interchangeability between states within the tristable parameter range, since the force of each attractor and corresponding barriers between the states are relatively low. Such transient tristability has previously been linked to cell fate decisions, and could represent a Transition State-promoting regime within the system (see Main Text). Importantly, the *SbP* also maintains discrete states throughout the range of parameter values (A, B and C) reiterating that the transition is not a gradual or continuous state separation, but comprised of discrete state-switching events.

Box 2 **| Single-cell Transcriptional Analysis Toolbox**

The growth in single-cell transcriptional profiling has required the parallel development of tools and methods to find patterns in such large datasets. Some of the existing single-cell transcriptional analysis tools are presented in Table 1.

Application	**Tools**	**Description**
Dimensionality Reduction Techniques	• Component Analysis; including Principle (PCA), Independent (ICA) and Multiresolution (MCA). • Multidimensional Scaling (MDS)[109] • t-Distributed Stochastic Neighbour Embedding (tSNE)[110] • Diffusion Maps[111]	Reduces multidimensional data to a minimal number of dimensions for visualisation, by identifying those dimensions that capture the important information in the data structure.
Gene Clustering	Self-Organising Maps (SOM)[112,113]	SOMs are unsupervised neural network learning algorithms that organise genes into biological relevant clusters which can then be compared between samples.
Clustering Methods	Various, including hierarchical, k-means, shared-nearest neighbour and many others.	A collection of methods that attempt to group observations based on similarity. With single-cell transcriptional data, these can be used to find populations of cells, or cohorts of genes.
Dynamic Clustering	Time-variant clustering[114]	Using statistical methods, cells can be spatially clustered at each time point and the relationship of clusters across time points described. Can be used to generate 'branching' patterns of cell clusters over time.
Trajectory Reconstruction	'Pseudotime' methods, with tools including Monocle[69] and Wanderlust[68].	Orders cells by progress through a dynamic process based on similarity and arranges them into a 'trajectory'.
Network Analysis	Gene Regulatory Network (GRN) Inference methods (see[89,115])	These methods seek to identify the underlying GRN that is responsible for the observed transcriptional patterns.

Figure Legends

Figure 1 **| Waddington's Epigenetic Landscape and modern representations. |A|** The classical view in which a cell, represented as a pebble, starts at the top of a hill and rolls down the landscape through a series of branching points that represent decision events. **|B|** Waddington conceived that the landscape was underpinned by the activity of genes, represented as pegs underneath the hills and valleys. **|C|** Cell fate decisions are coordinated through the regulatory interactions between genes (***x and y***). Simple network motifs such as the bistable switch (**left**) can be modelled, and phase-planes drawn, which identify the stable (***filled red circles***) and metastable (***empty purple circle***) points of the system (**middle**). With this framework, quasipotential values for each potential state of the system can be calculated and plotted as a third dimension (**right**). **|D|** By invoking a time-dependent parameter of the network (***lamda***), the dynamic response of the system can be examined. Such parameter changes represent biological situations in which cellular signalling inputs alter the network and affect fate decisions (**left**). At critical parameter values, the topology of the system can change, for instance by converting a monostable system with one stable point (***red line***) into a bistable one with two stable (***red lines***) and one metastable points (***purple dotted line***; **middle**). Calculating the quasipotential values in this system can give a dynamic landscape (**right**), highly reminiscent of Waddington's original Epigenetic landscape. Figure 1 A & B were reproduced with permission from Waddington (1957) *The Strategy of the Genes*[14].

Figure 2 **| Continuous and Discrete analysis of Cell Fate Decisions from Single Cell Gene Expression data. |A|** In a fate transition, a cell may traverse the transcriptional phase space in a smooth, continuous manner before reaching a decision point. Single-cell transcriptional data from cell ensembles can provide a snapshot of the population structure, and transcriptional profiles of each cell can be arranged into hierarchical clusters of varying similarity; the data might also contain information about the dynamics and mechanisms of the decision event. There are two alternative interpretations of this information: **|B|** If the transition between states is assumed to be continuous, cells are ordered along a sequence that assumes that more similar cells should be closer together on the trajectory of differentiation; a sequence often referred to as 'pseudotime'. **|C|** Instead, if the transition between states is deemed to be discontinuous, as in the case of the Transition State, cells might cluster into a number of functional 'macrostates', each with a number of corresponding 'microstates' (see Main Text for details). Each of these microstates has a different probability of transitioning to another state, but this is not necessarily directly correlated to similarities between a given microstate and the final state. While there may be continuous processes leading up to and after decision events, the decision itself is *discontinuous* and stochastic*,* where each state has an associated *probability* of transitioning to any other state. **|D-E|** Experimental evidence of a broad transcriptional space in haematopoietic cells undergoing commitment decisions. Principal component (PC) analysis plots of single-cell RT-qPCR data for cultured (D, EML) and primary (E, mouse bone marrow, BM) cells undergoing erythroid commitment and differentiation decisions. Plots are a comprehensive representation of the data in Pina *et. al.*[81], and highlight the point that early committed cells (ECP, in the case of EML; preMegE, in the case of primary BM) are more heterogeneous in their transcriptional programmes than the multipotent self-renewing cells (EML, SR; BM, long-term and short-term HSC) they originate from, or the differentiated progeny (EML, Ediff; BM, CFUe) they give rise to. A similar observation was made by Jaenisch and collaborators[77] in analysing transcriptional programmes of individual cells undergoing early fate transitions in iPS reprogramming. SR: self-renewing cells; ECP - erythroid-committed population obtained in two distinct cytokine conditions, SCF: stem cell factor, Epo: erythropoietin; Ediff: differentiated erythroid cells. LT-HSC: Long-term reconstituting haematopoietic stem cell; ST-HSC: Short-term reconstituting HSC; preMegE: pre-megakaryocytic/erythroid progenitor; CFUe: colony-forming unit-erythroid (see Pina et al NCB 2012[81] for details).

Figure 3 **| The Transition State. |A|** In Chemistry, a Transition State (TS) is a short-lived intermediate that arises in a chemical reaction, in which reactant atoms assume a configuration that is intermediate to the initial and final products and which explores an available space of potential energy. In endergonic reactions, the free energy of the initial state must overcome an Activation Energy before assuming the lower-energy final state with the products of the reaction. Additional components such as catalysts can alter the height of the activation energy, changing the likelihood of transition. **|B|** By analogy, biological gene regulatory networks could be deemed to undergo similar transition dynamics as they make cell fate decisions. The activation of a specific node within a gene regulatory network leads the system to assume a TS, in which a number of network states are represented that are intermediate to either the initial or final states of the network. Each has identical network structure, but with different gene expression levels (represented by node colour), and with an associated probability of transitioning to any other state, including a transition back to the original state. We surmise that the TS represents a heterogeneous substrate for network selection and fate decisions by factors such as signalling pathways.

References

1. Sulston, J.E. & Horvitz, H.R. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol* **56**, 110-56 (1977).
2. Nishida, H. Specification of embryonic axis and mosaic development in ascidians. *Dev Dyn* **233**, 1177-93 (2005).
3. Davidson, E.H. *The regulatory genome: gene regulatory networks in development and evolution*, (Academic press, 2010).
4. Levine, M. & Davidson, E.H. Gene regulatory networks for development. *Proc Natl Acad Sci U S A* **102**, 4936-42 (2005).
5. Mathis, L. & Nicolas, J.F. Cellular patterning of the vertebrate embryo. *Trends Genet* **18**, 627-35 (2002).
6. Davidson, E.H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669-78 (2002).
7. Stathopoulos, A. & Levine, M. Genomic regulatory networks and animal development. *Dev Cell* **9**, 449-62 (2005).
8. Kamminga, L.M. *et al.* Autonomous behavior of hematopoietic stem cells. *Exp Hematol* **28**, 1451-9 (2000).
9. Luer, K. & Technau, G.M. Single cell cultures of Drosophila neuroectodermal and mesectodermal central nervous system progenitors reveal different degrees of developmental autonomy. *Neural Dev* **4**, 30 (2009).
10. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-55 (2005).
11. Loebel, D.A.F., Watson, C.M., De Young, R.A. & Tam, P.P.L. Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Developmental Biology* **264**, 1-14 (2003).
12. Hayward, P., Kalmar, T. & Martinez-Arias, A. Wnt/Notch signalling and information processing during development. *Development* **135**, 411-24 (2008).
13. Waddington, C.H. Canalization of development and the inheritance of acquired characteristics. *Nature* **3811**, 563-565 (1942).
14. Waddington, C.H. *The strategy of the genes: a discussion of some aspects of theoretical biology*, (Allen & Unwin, 1957).
15. Allen, M. Compelled by the Diagram: Thinking through C. H. Waddington’s Epigenetic Landscape. *2015* **4**, 24 (2015).
16. Kauffman, S. Homeostasis and differentiation in random genetic control networks. *Nature* **224**, 177-8 (1969).
17. Kauffman, S.A. *The Origins of Order: Self Organization and Selection in Evolution*, (Oxford University Press, 1993).
18. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D.E. Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Physical Review Letters* 94, 128701 (2005).
    An exposition of Waddington landscape in terms of dynamical systems theory.
19. Huang, S. The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? *BioEssays* **34**, 149-157 (2012).
    A clear discussion of the history and meaning of Waddington's landscape.
20. Wang, J., Xu, L., Wang, E. & Huang, S. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys J* **99**, 29-39 (2010).
21. Trott, J., Hayashi, K., Surani, A., Babu, M.M. & Martinez-Arias, A. Dissecting ensemble networks in ES cell populations reveals micro-heterogeneity underlying pluripotency. *Mol Biosyst* **8**, 744-52 (2012).
    A discrete, discontinuous view of a cell state from the perspective of statistical mechanics.
22. Marr, C., Zhou, J.X. & Huang, S. Single-Cell Gene Expression Profiling and Cell State Dynamics: Collecting Data, Correlating Data Points and Connecting the Dots. *bioRxiv* (2016).

23. Jaeger, J., Manu & Reinitz, J. Drosophila blastoderm patterning. *Curr Opin Genet Dev* **22**, 533-41 (2012).
24. Ingham, P.W. The molecular genetics of embryonic pattern formation in Drosophila. *Nature* **335**, 25-34 (1988).
25. Ferguson, E.L., Sternberg, P.W. & Horvitz, H.R. A genetic pathway for the specification of the vulval cell lineages of Caenorhabditis elegans. *Nature* **326**, 259-67 (1987).
26. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*, (Taylor & Francis, 2006).
27. Vermeirssen, V. *et al.* Transcription factor modularity in a gene-centered C. elegans core neuronal protein–DNA interaction network. *Genome Research* **17**, 1061-1071 (2007).
28. Arda, H.E. *et al.* Functional modularity of nuclear hormone receptors in a Caenorhabditis elegans metabolic gene regulatory network. *Mol Syst Biol* **6**, 367 (2010).
29. MacNeil, L.T. & Walhout, A.J.M. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research* **21**, 645-657 (2011).
30. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824-7 (2002).
31. Shen-Orr, S.S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet* **31**, 64-8 (2002).
32. Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-61 (2007).
33. Edgar, B.A., Odell, G.M. & Schubiger, G. A genetic switch, based on negative regulation, sharpens stripes in Drosophila embryos. *Dev Genet* **10**, 124-42 (1989).
34. Wang, L. *et al.* Bistable switches control memory and plasticity in cellular differentiation. *Proc Natl Acad Sci U S A* **106**, 6638-43 (2009).
35. Bouldin, C.M. *et al.* Wnt signaling and tbx16 form a bistable switch to commit bipotential progenitors to mesoderm. *Development* **142**, 2499-507 (2015).
36. Bhattacharya, S., Zhang, Q. & Andersen, M.E. A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst Biol* **5**, 85 (2011).
37. Verd, B., Crombach, A. & Jaeger, J. Classification of transient behaviours in a time-dependent toggle switch model. *BMC Syst Biol* **8**, 43 (2014).

    An important consideration of the role that temporal changes in network parameters has on Waddington's landscape.
38. Huang, S., Guo, Y.P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* **305**, 695-713 (2007).

    An important, concrete and clear discussion of a fate choices in terms of dynamical systems theory.
39. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-26 (2008).
40. Eldar, A. & Elowitz, M.B. Functional roles for noise in genetic circuits. *Nature* **467**, 167-73 (2010).
41. Schröter, C., Rué, P., Mackenzie, J.P. & Martinez-Arias, A. FGF/MAPK signaling sets the switching threshold of a mutual repressor circuit controlling cell fate decisions in ES cells. *bioRxiv* (2015).
42. Süel, G.M., Garcia-Ojalvo, J., Liberman, L.M. & Elowitz, M.B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545-550 (2006).
43. Kalmar, T. *et al.* Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* **7**, e1000149 (2009).
44. Martinez-Arias, A. & Brickman, J.M. Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Current Opinion in Cell Biology* **23**, 650-656 (2011).
45. Martinez-Arias, A. & Hayward, P. Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet* **7**, 34-44 (2006).

46. Ahrends, R. *et al.* Controlling low rates of cell differentiation through noise and ultrahigh feedback. *Science (New York, N.Y.)* **344**, 1384-1389 (2014).
47. Grun, D. & van Oudenaarden, A. Design and Analysis of Single-Cell Sequencing Experiments. *Cell* **163**, 799-810 (2015).
48. Stegle, O., Teichmann, S.A. & Marioni, J.C. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* **16**, 133-45 (2015).
49. Bendall, S.C., Nolan, G.P., Roederer, M. & Chattopadhyay, P.K. A deep profiler's guide to cytometry. *Trends Immunol* **33**, 323-32 (2012).
50. Jaitin, D.A., Keren-Shaul, H., Elefant, N. & Amit, I. Each cell counts: Hematopoiesis and immunity research in the era of single cell genomics. *Semin Immunol* **27**, 67-71 (2015).
51. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res* **25**, 1491-8 (2015).
52. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat Meth* **10**, 1093-1095 (2013).
53. Grun, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature Methods* **11**, 637 - 40 (2014).
54. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-4 (2007).
55. Singer, Z.S. *et al.* Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells. *Molecular Cell* **55**, 319-331 (2014).
56. Abranches, E. *et al.* Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* **141**, 2770-9 (2014).
57. Grun, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251-5 (2015).
58. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotech* **33**, 155-160 (2015).
59. Iwasaki, H. & Akashi, K. Hematopoietic developmental pathways: on cellular basis. *Oncogene* **26**, 6687-6696 (0000).
60. Doulatov, S., Notta, F., Laurenti, E. & Dick, John E. Hematopoiesis: A Human Perspective. *Cell Stem Cell* **10**, 120-136 (2012).
61. Drissen, R. *et al.* Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol* (2016).
62. Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663-77 (2015).
63. Takano, H., Ema, H., Sudo, K. & Nakauchi, H. Asymmetric division and lineage commitment at the level of hematopoietic stem cells: inference from differentiation in daughter cell and granddaughter cell pairs. *J Exp Med* **199**, 295-302 (2004).
64. Pronk, C.J. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428-42 (2007).
65. Arinobu, Y. *et al.* Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell* **1**, 416-27 (2007).
66. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I.L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-7 (2000).
67. Teles, J. *et al.* Transcriptional Regulation of Lineage Commitment - A Stochastic Model of Cell Fate Decisions. *PLoS Comput Biol* **9**, e1003197 (2013).

How to infer regulatory relationships from single cell transcriptional data.

68. Bendall, Sean C. *et al.* Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. *Cell* **157**, 714-725.
69. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotech* **32**, 381-386 (2014).

70. Moignard, V. *et al.* Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat Biotech* **33**, 269-276 (2015).

One of the first accounts of fate transitions in embryos at the level of single cells.

71. Ocone, A., Haghverdi, L., Mueller, N.S. & Theis, F.J. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* **31**, i89-i96 (2015).
72. Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* **11**, 774-85 (1997).

Pioneering work on lineage priming and transcriptional heterogeneities in single cells.

73. Goolam, M. *et al.* Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell* **165**, 61-74 (2016).
74. Brunskill, E.W. *et al.* Single cell dissection of early kidney development: multilineage priming. *Development* **141**, 3093-101 (2014).
75. Miyamoto, T. *et al.* Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev Cell* **3**, 137-47 (2002).
76. Laslo, P. *et al.* Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* **126**, 755-66 (2006).
77. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209-22 (2012).
78. Piras, V., Tomita, M. & Selvarajoo, K. Transcriptome-wide variability in single embryonic development cells. *Sci Rep* **4**, 7137 (2014).
79. Nair, G., Abranches, E., Guedes, A.M., Henrique, D. & Raj, A. Heterogeneous lineage marker expression in naive embryonic stem cells is mostly due to spontaneous differentiation. *Sci Rep* **5**, 13339 (2015).
80. Kumar, R.M. *et al.* Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516**, 56-U112 (2014).
81. Pina, C. *et al.* Inferring rules of lineage commitment in haematopoiesis. *Nat Cell Biol* **14**, 287-94 (2012).
82. Muñoz Descalzo, S., RuÉ, P., Garcia-Ojalvo, J. & Martinez-Arias, A. Correlations Between the Levels of Oct4 and Nanog as a Signature for Naïve Pluripotency in Mouse Embryonic Stem Cells. *STEM CELLS* **30**, 2683-2691 (2012).
83. Garcia-Ojalvo, J. & Martinez-Arias, A. Towards a statistical mechanics of cell fate decisions. *Current Opinion in Genetics & Development* **22**, 619-626 (2012).
84. Cross, M.A. & Enver, T. The lineage commitment of haemopoietic progenitor cells. *Curr Opin Genet Dev* **7**, 609-13 (1997).
85. Mitschka, S. *et al.* Co-existence of intact stemness and priming of neural differentiation programs in mES cells lacking Trim71. *Scientific Reports* **5**, 11126 (2015).
86. Munoz-Descalzo, S., de Navascues, J. & Martinez-Arias, A. Wnt-Notch signalling: an integrated mechanism regulating transitions between cell states. *Bioessays* **34**, 110-8 (2012).
87. Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-7 (2008).
88. Kutejova, E., Sasai, N., Shah, A., Gouti, M. & Briscoe, J. Neural Progenitors Adopt Specific Identities by Directly Repressing All Alternative Progenitor Transcriptional Programs. *Dev Cell* **36**, 639-53 (2016).
89. Pina, C. *et al.* Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. *Cell Reports* **11**, 1503-1510.
90. Laidler, K.J. & King, M.C. Development of transition-state theory. *The Journal of Physical Chemistry* **87**, 2657-2664 (1983).
91. Trott, J. & Martinez-Arias, A. Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. *Biol Open* **2**, 1049-56 (2013).

92. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**(2016).

93. Turner, D.A. *et al.* Wnt/beta-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development* **141**, 4243-53 (2014).

94. Ferrell, J.E., Jr. Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr Biol* **22**, R458-66 (2012).

Considers how to incorporate time and irreversibility in epigenetic landscapes.

95. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369**(2014).

96. Nichols, J. & Smith, A. Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect Biol* **4**, a008128 (2012).

97. Rue, P. & Martinez-Arias, A. Cell dynamics and gene expression control in tissue homeostasis and development. *Mol Syst Biol* **11**, 792 (2015).

98. Balazsi, G., van Oudenaarden, A. & Collins, J.J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910-25 (2011).

99. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183-1186 (2002).

100. Swain, P.S., Elowitz, M.B. & Siggia, E.D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* **99**, 12795-800 (2002).

101. Morikawa, M., Koinuma, D., Miyazono, K. & Heldin, C.H. Genome-wide mechanisms of Smad binding. *Oncogene* **32**, 1609-15 (2013).

102. Schmierer, B. & Hill, C.S. TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol* **8**, 970-82 (2007).

103. Ohnishi, Y. *et al.* Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat Cell Biol* **16**, 27-37 (2014).

104. Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. & Hadjantonakis, A.K. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081-91 (2008).

105. Frankenberg, S. *et al.* Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev Cell* **21**, 1005-13 (2011).

106. Bessonnard, S. *et al.* Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**, 3637-48 (2014).

107. De Mot, L. *et al.* Cell Fate Specification Based on Tristability in the Inner Cell Mass of Mouse Blastocysts. *Biophys J* **110**, 710-22 (2016).

108. Strogatz, S.H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, (Westview Press, 1994).

109. Cox, A.M.A. & Cox, F.T. Multidimensional Scaling. in *Handbook of Data Visualization* 315-347 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).

110. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 85 (2008).

111. Coifman, R.R. & Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis* **21**, 5-30 (2006).

112. Tamayo, P. *et al.* Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences* **96**, 2907-2912 (1999).

113. Törönen, P., Kolehmainen, M., Wong, G. & Castrén, E. Analysis of gene expression data using self-organizing maps. *FEBS Letters* **451**, 142-146 (1999).

114. Huang, W., Cao, X., Biase, F.H., Yu, P. & Zhong, S. Time-variant clustering model for understanding cell fate decisions. *Proceedings of the National Academy of Sciences* **111**, E4797-E4806 (2014).

115. Park, J., Ogunnaike, B., Schwaber, J. & Vadigepalli, R. Identifying functional gene regulatory network phenotypes underlying single cell transcriptional variability. *Prog Biophys Mol Biol* **117**, 87-98 (2015).

Online Summary

- During cell fate decisions, cells go through binary fate choices, fuelled by the activity of Gene Regulatory Networks (GRNs): integrated systems of interactions between transcription factors and signal transduction effectors.
- The Waddington landscape is an abstract illustration of cell fate dynamics which aims to capture the activity of GRNs in cell fate decision. Formal representations of the landscape represent fate decisions as smooth, continuous events in which cells follow predetermined trajectories.
- Advances in single-cell transcriptional methods and their associated analytical tools can potentially reveal the dynamics of transition events. These experiments are revealing a high degree of transcriptional heterogeneity particularly at decision points.
- Single cell transcriptomes are used to reveal the structure of populations but they might also contain information about the mechanisms underpinning fates and fate transitions. For example, 'pseudotime' algorithms use high-dimensional data to order cells by similarity along a continuum which can be easily mapped onto the continuum of the classical Waddington landscape.
- An alternative interpretation of transcriptional heterogeneities, particularly at the fate transition states, lies in the notion of 'Transition state' (TS), a transient intermediate characterized by an ensemble of many possible network profiles associated with the states available to individual cells.
- From the perspective of the TS, transcriptional heterogeneities represent ensembles of unstable, discrete networks that relate the initial and final stages: fate decisions are discontinuous events.
- Signalling and other factors bias decisions by favouring particular states in the TS, by using the TS as a heterogeneous substrate for cellular selection, or act to change the probability of transition associated with each configuration. As such, the system can be tuned to regulate the size of a given population and can therefore provide a tractable concept for cell fate decisions in development and homeostasis.

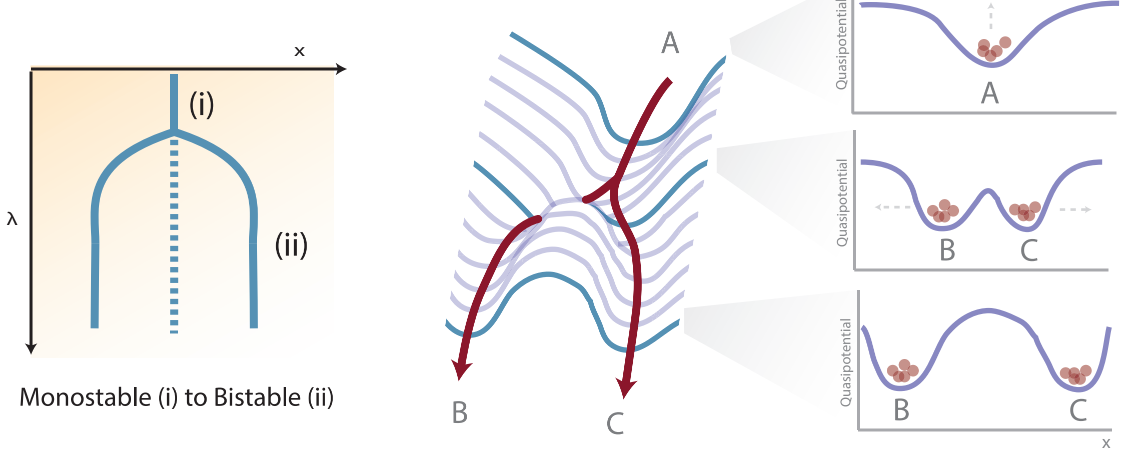
Author Biographies

Naomi Moris is completing her PhD at the University of Cambridge where she studies cell fate decisions in embryonic stem cells, particularly focussing on the role of chromatin modifiers in regulating transcriptional heterogeneity and its implications for decision-making in single cells.

Alfonso Martinez Arias has been at the University of Cambridge for 30 years, first as a Wellcome Trust Fellow in the Department of Zoology and later as a Professor of Developmental Mechanics in the Department of Genetics. His laboratory studies mechanisms of cell fate decisions in animal development.

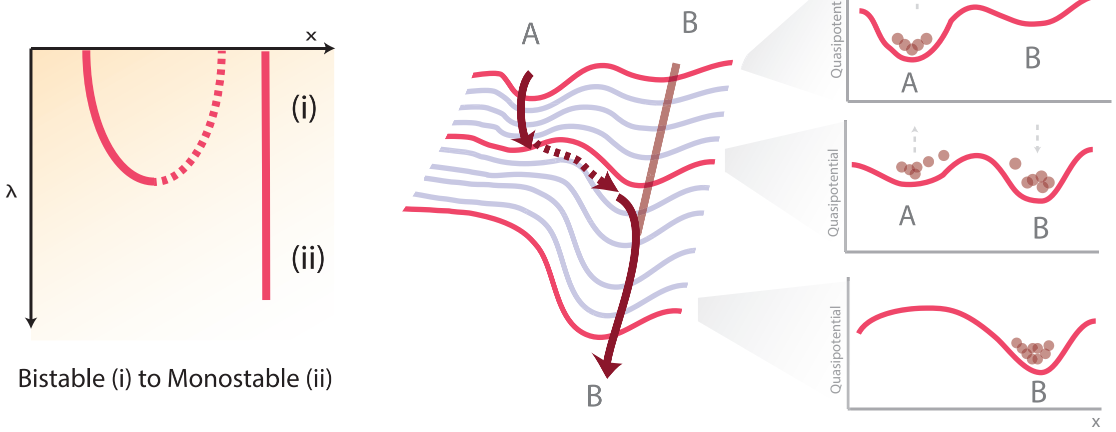
Cristina Pina is a KKLF Intermediate Fellow and Principal Investigator at the University of Cambridge Department of Haematology. Medically-trained, she studied transcriptional regulation of haematopoietic lineage commitment with Tariq Enver in Oxford and at UCL. Her research focuses on the role and regulation of transcriptional noise in malignant fate decisions.

Supercritical Pitchfork Bifurcation



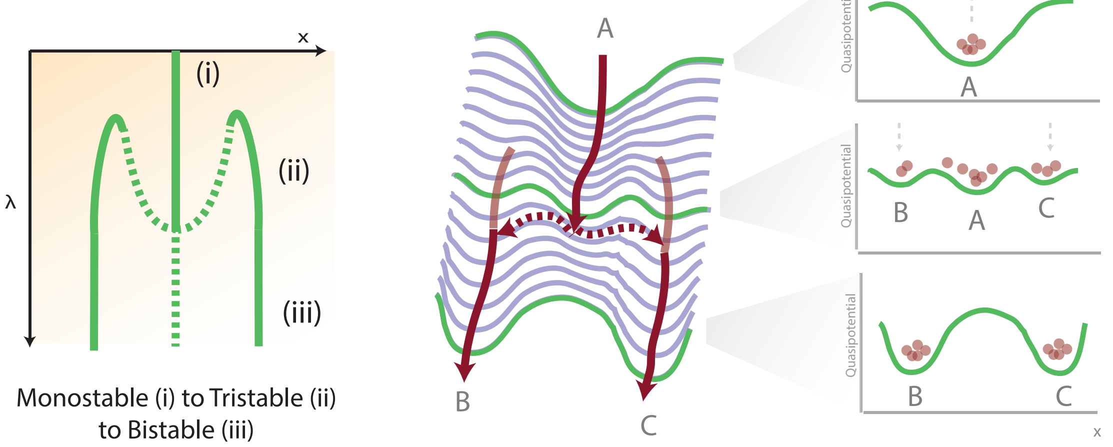
Monostable (i) to Bistable (ii)

Saddle-node Bifurcation

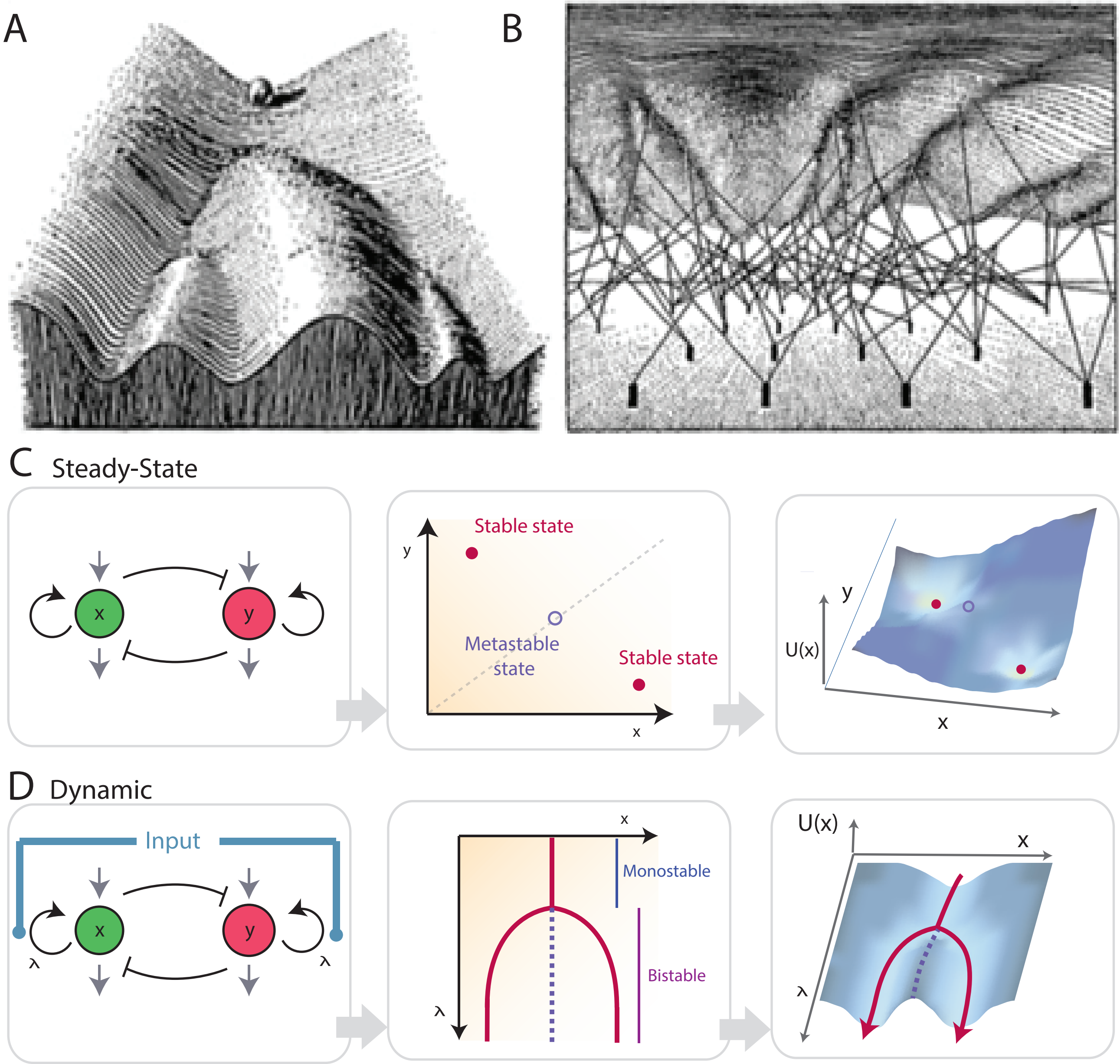


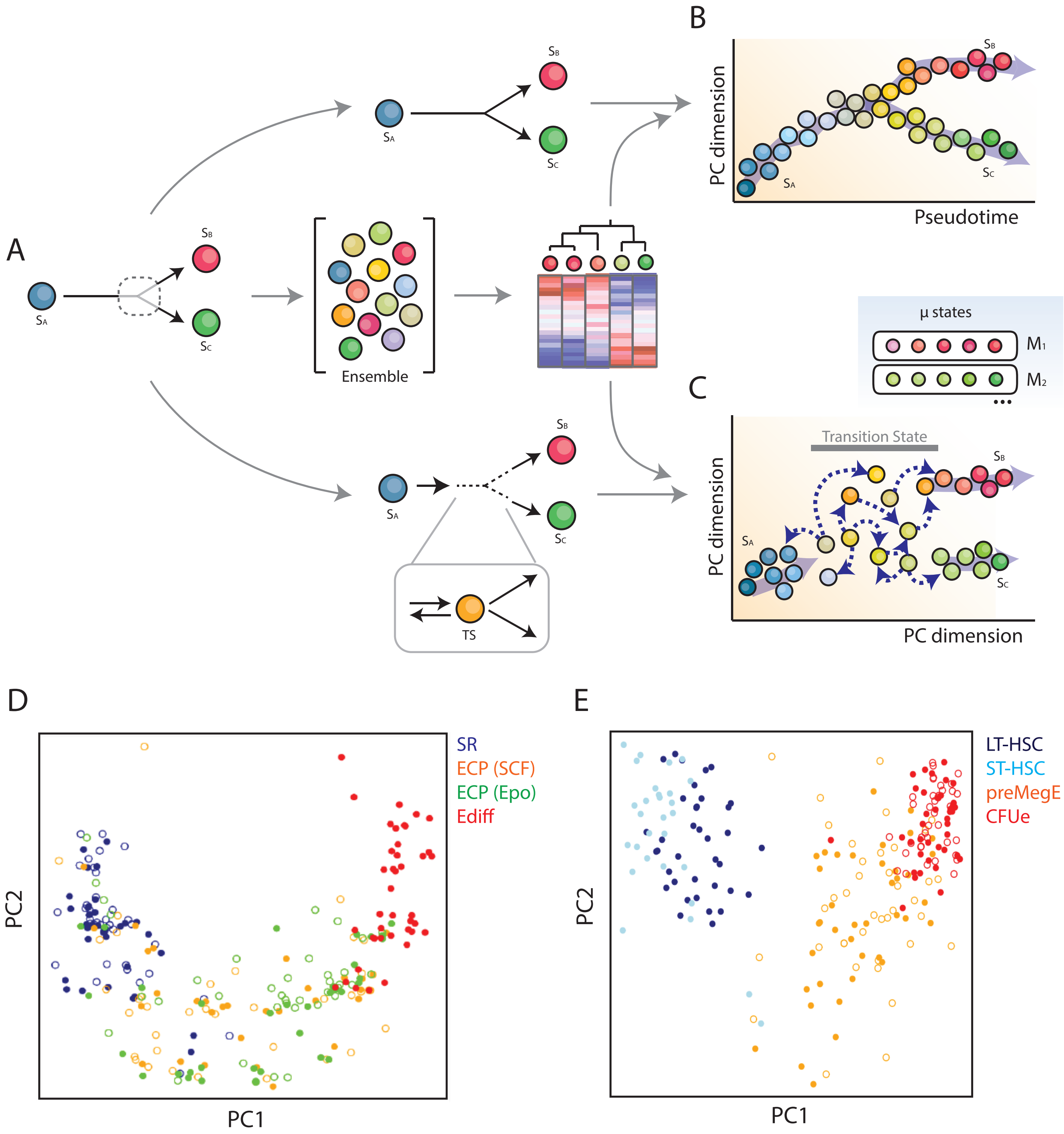
Bistable (i) to Monostable (ii)

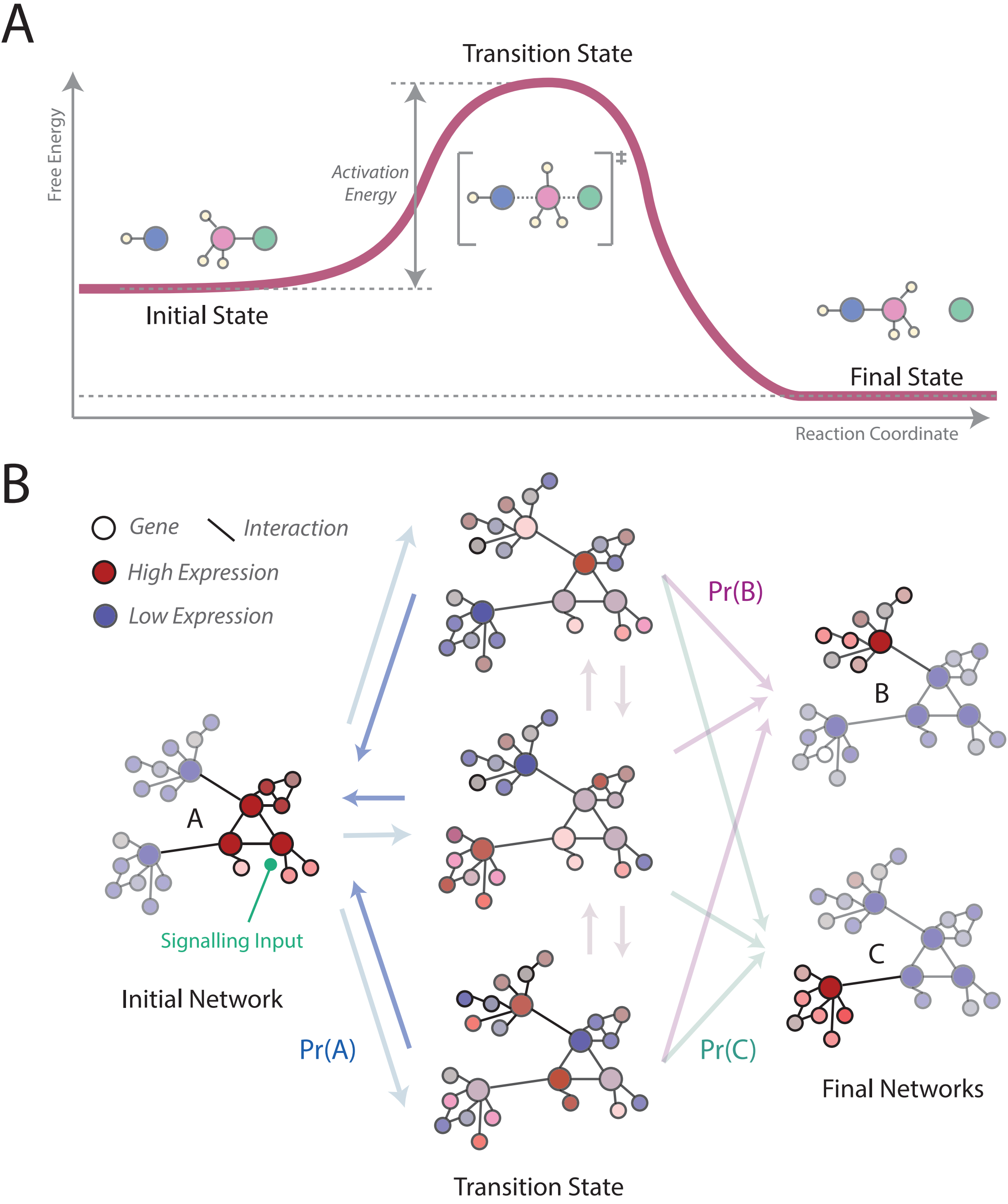
Subcritical Pitchfork Bifurcation



Monostable (i) to Tristable (ii)
to Bistable (iii)


A
B
C Steady-State
x
y
Stable state
y
Metastable state
Stable state
x
y
U(x)
x
D Dynamic
Input
x
y
λ
λ
x
Monostable
Bistable
λ
U(x)
x
λ


A
S_A
S_B
S_C
Ensemble
TS
B
PC dimension
Pseudotime
μ states
M_1
M_2
...
C
Transition State
PC dimension
PC dimension
D
PC2
PC1
SR
ECP (SCF)
ECP (Epo)
Ediff
E
PC2
PC1
LT-HSC
ST-HSC
preMegE
CFUe


A
Transition State
Free Energy
Activation Energy
‡
Initial State
Final State
Reaction Coordinate
B
Gene
Interaction
High Expression
Low Expression
Pr(B)
B
A
Signalling Input
Initial Network
Pr(A)
Pr(C)
C
Final Networks
Transition State