Universal principles of lineage architecture and stem cell identity in renewing tissues

Philip Greulich[1, 2,†], **Ben D. MacArthur**[1, 2, 3], **Cristina Parigini**[1, 2], **and Rubén J. Sánchez-García**[1, 2]

[1]Mathematical Sciences, University of Southampton, Highfield Campus, Southampton SO17 1BJ, United Kingdom; [2]Institute for Life Sciences, University of Southampton, Highfield Campus, Southampton SO17 1BJ, United Kingdom; [3]Centre for Human Development, Stem Cells and Regeneration, Faculty of Medicine, University of Southampton, Southampton SO17 1BJ, United Kingdom; [†]Corresponding author

Adult tissues in multicellular organisms typically contain a variety of stem, progenitor and differentiated cell types arranged in a lineage hierarchy that regulates healthy tissue turnover. Lineage hierarchies in disparate tissues often exhibit common features, yet the general principles regulating their architecture are not known. Here, we provide a formal framework for understanding the relationship between cell molecular 'states' and cell 'types', based on the topology of admissible cell state trajectories. We show that a self-renewing cell type – if defined as suggested by this framework – must reside at the top of any homeostatic renewing lineage hierarchy, and only there. This architecture arises as a natural consequence of homeostasis, and indeed is the only possible way that lineage architectures can be constructed to support homeostasis in renewing tissues. Furthermore, under suitable feedback regulation, for example from the stem cell niche, we show that the property of 'stemness' is entirely determined by the cell environment, in accordance with the notion that stem cell identities are contextual and not determined by hard-wired, cell-intrinsic, characteristics.

1. Introduction

Adult stem cells were first identified in the 1960s context of hematopoiesis (1, 2) and are now believed to reside in all renewing tissues. Because of their ubiquity and regenerative importance, the quest to find a complete set of defining characteristics of adult stem cells has been a longstanding aim of stem cell and developmental biology. However, despite numerous attempts, we still lack an unambiguous characterisation, and the stem cell concept is still widely disputed (3–8).

Historically, the stem cell identity has been viewed as a hard-wired phenotype, which is encoded in distinct patterns of gene and protein expression that confer to the cell the ability to maintain its own state while regenerating the cells of a whole tissue. Modern experimental methods have, however, cast doubt on this interpretation. For example, single-cell transcriptomics experiments have revealed a greater degree of molecular heterogeneity within apparently functionally 'pure' cell populations (including stem cell populations) than previously envisaged, and the search for unique stem cell markers has often, accordingly, proven challenging (9, 10). Perhaps more significantly, genetic cell lineage tracing studies have shown that many cells can exhibit a high degree of functional plasticity. For instance, cells which under homeostatic conditions are committed to differentiation are, in other contexts, such as regeneration or cancer, apparently able to acquire stem cell associated features, and can re-populate the stem cell pool via de-differentiation (11–16). In response to these studies it has been suggested that rather than being a hard-wired cell-intrinsic property, 'stemness' – i.e. the attribute of being a stem cell – is a functional, dynamical cellular feature that depends on the cellular environment (3, 6, 17).

In practice, the most commonly accepted characterisation of stem cells is purely pragmatic: an adult stem cell is a cell which has the capability to (i) maintain its own population via self-renewing cell divisions and (ii) produce all cells of a specific tissue lineage via differentiation. According to this functional characterisation, 'stemness' is assessed by a cell's lineage potential – i.e. what a stem cell and its progeny can become in the future, rather than by its molecular characteristics at present.

Here, we will take this basic functional definition as a starting point and seek to examine the relationships between intracellular molecular 'states' and cell 'types' in order to clarify, precisely, what we mean when we refer to a stem cell. We start by acknowledging that a cell may be characterised by its internal molecular state (which we will define precisely shortly), and the possible changes in cell states which confer changes in cell function, for example when differentiating. It is reasonable, therefore, to define a cell both in terms of its current state, and the states that it may reach in the future under appropriate stimulus: that is, by the set of admissible cell state trajectories.

Notably, such cell state trajectories may not just be simple paths: they can have a complex topology, including branching points at which cells make fate 'decisions' and closed loops or cycles (corresponding to recurrent dynamics, such as the cell cycle). Our essential insight is that we can represent all possible cell states and changes between them as a directed network.

Cell state trajectories are then directed paths on this network. Importantly, the structure of this network defines which trajectories are allowed and which are not and so is of central importance.

Here we analyse the general structural and dynamical properties of the network of cell state trajectories, and use this analysis to precisely define stem cell identities and cell lineages in adult renewing tissues. Typically, we have very limited experimental information available about this potentially very complex network of cell states and transitions. However, we will show, using notions from dynamical systems and network theory, that homeostasis imposes dynamical constraints that ensure that the lineage hierarchy of adult renewing tissues is necessarily very constrained.

The key finding of this article is that functional definitions of 'stem' and committed cells arise naturally within this framework. Particularly, we show that there is an inherent relationship between the classification of self-renewing versus committed cell types and their position in the cell lineage hierarchy of homeostatic tissues: any self-renewing cell type, if defined as suggested, must reside at an apex of a homeostatic cell lineage. In consequence, self-renewing cells always have maximal lineage potential in homeostatic tissues and can therefore be classified as stem cells. Thus, self-renewal and lineage potential – the two defining characteristics of stem cells – are inherently connected.

While this is commonly experimentally observed, there is currently no rationale for why this is the case. We demonstrate that this commonly observed structural hierarchy is not a mere coincidence of evolution, but instead, is the only possible lineage architecture that is able to maintain homeostasis in renewing tissues, in any complex, multi-cellular, life form. Within this framework, it also becomes apparent why stemness is not necessarily a cell-intrinsic property, but rather is an emergent property of populations of cells interacting with their cellular environment, for example via crowding regulation or instruction from a niche.

2. Results

Many adult tissues, such as the epidermis, intestinal epithelium and other epithelia, are constantly renewing: old cells are steadily removed and new cells are produced by cell division. In undisturbed conditions (i.e. in the absence of disease or trauma), adult tissue renewal is in homeostasis – the state of a tissue that is fully developed, unstressed, and behaving physiologically. In particular, during homeostatic tissue renewal, the tissue cell composition does not change significantly over time and this necessarily requires cell division, differentiation and death to be finely balanced. Our aim is to understand the constraints imposed by homeostasis on the structure of possible cell lineage architectures.

Typically, not all cells in a tissue can divide. Instead, only a fraction have the ability to keep dividing in the long-term, with their progeny differentiating to provide a constant turnover of tissue-specific cells. It has been disputed whether such *self-renewing* progenitor cells generally qualify as 'adult stem cells' (4, 5, 8). Here, we show that constraints on tissue renewal imposed by homeostasis are sufficient to unambiguously define the stem cell notion in a mathematically robust and coherent way in any homeostatic renewing tissue.

A. Cell types and cell lineages.

A.1. Cell state trajectories. To start, we assume that a cell's phenotype can be characterised by its internal molecular composition, including expression of RNAs, proteins, and metabolic components, etc. We will be interested not only in the current state of the cell, but also its potential, including its propensity to divide, change its phenotypic state, die or emigrate out of the tissue of interest. Rather than restrict to any specific technology (e.g. single-cell RNA sequencing) we define a cell's state in terms of the molecular features (measurable or not) which determine these propensities.

The molecular composition of a cell will naturally change over time, but not all molecular changes will be possible at any instance in time. We will refer to any admissible transition between cell states – a directed path from an initial state to a final state – as a *cell state trajectory*. The set of all admissible cell state trajectories determines a cell's lineage potential. Notably, the set of admissible cell state trajectories can have a complex topology: it can contain branching points, for example, at which trajectories diverge to different cell fates. Moreover, trajectories may be reversible and form cycles (for example, the cell cycle). Crucially, we can use the topology of the set of cell state trajectories as a basis to define cell types, lineages, and stem cell identity.

P.G., B.D.M. and R.S.G conceived the project, P.G., B.D.M. and R.S.G wrote the paper, P.G. supervised the research, C.P. carried out the numerical analysis, and P.G. and R.S.G. carried out the mathematical analysis.

The authors declare no conflict of interest.

[2] Correspondence should be addressed to P.G., E-mail: P.S.Greulich@soton.ac.uk

A.2. Cell types. Cell types are typically defined by molecular characteristics, for example expression of surface markers or via clustering from single-cell transcriptome data. Although pragmatic, such taxonomies do not provide a rationale for why certain molecular states are associated with specific cell functions. However, we can assess a cell's lineage potential by analysing the set of possible cell state trajectories to which it belongs and are thus able to classify cell states according to the role they play in tissue development and maintenance via their position in the lineage hierarchy.

To define a cell type in this context, we propose that cells of the same type should have the same lineage potential, and vice versa. Thus, any two states i and j belonging to the same cell type should share the same outgoing cell state trajectories. This is the case if and only if states i and j are able to interconvert, that is, if they are mutually reachable via cell state trajectories. In this case, any state reachable from i is also reachable from j. According to this rationale we can formulate a definition of a cell type:

Definition 1. *A* cell type *is a maximal set of cell states which are mutually reachable by cell state trajectories.*

Informally, this means that we can decompose the set of all possible cell state trajectories into disjoint groups, such that all cell states in the same group are connected via cyclic cell state trajectories (through which they are mutually reachable), and no pair of cells in different groups are connected by cyclic trajectories. In this formulation, each such group constitutes a cell type.

We note that this definition of cell type is not always congruent with informal yet commonly used definitions of a cell type. Typically, cell types are associated with distinct patterns of expression, or distinct morphologies or functions. In our alternative view, cells of the same type may have substantially different molecular profiles or morphologies, so long as they are interchangeable (i.e. spontaneous conversion from one state or morphology to the other is possible). In this view, instantaneous molecular or morphological similarity is not essential, rather it is the potential of the cell that matters – and so allows for the possibility that cell types can be inherently dynamic. To illustrate this notion, consider the cell cycle: during the cell cycle, a cell may change its molecular state and morphology dramatically, yet we would not consider each stage of the cell cycle as a different cell type. Consistent with this view, our definition groups all cell states within the cell cycle to one cell type.

Furthermore, we note that cell state trajectories, and therefore cell types, may vary with environmental context. For example, two distinct cell types may be present in one environment (e.g. *in vitro* or in grafts), yet in another environment (e.g. *in vivo*) a reversible transition between the states of those types may become possible, and thus, in our framework, only one cell type is present.

A.3. Cell state networks and lineages. In the following sections we formalise these arguments, using notions from network theory, and use this reasoning to propose universal features of lineage architectures under homeostatic conditions. To that end, we first represent cell trajectories as transitions between a discrete set of cell states. Although we take a discrete formalism, our results are also valid in a continuous conceptualisation since continuous trajectories can always be discretized in a topology-preserving way (18).

The set of discrete cell states and admissible cell state transitions forms a network, which we will call the *cell state network*. The nodes of this network are the cell states, and there is a directed link from node i to node j whenever a transition from cell state i to cell state j is possible without intermediate states (according to the chosen discretization). Directed paths (i.e. sequences of directed links) in this network represent allowed cell state trajectories. An example of such a cell state network is shown in Fig. 1, left.

Our intuitive definition of a cell type has a well-characterised counterpart in the terminology of network theory. Two nodes i, j that are mutually reachable via directed paths in a network are said to be *strongly connected*. Furthermore, any directed network may be uniquely decomposed into *strongly connected components (SCCs)* by grouping together all strongly connected nodes (19) (see Fig. 1 for an example). From a biological viewpoint, it is reasonable to consider only *non-trivial* SCCs, i.e. those that contain at least one link, (that is, either more than one node or one node with a self-link). Thus, our intuitive definition of a cell type may be naturally phrased in the language of network theory: a *cell type* is a non-trivial strongly connected component of the underlying cell state network.

Using this definition, we can construct a second network, with cell types (that is, SCCs of the cell state network) as nodes, and a directed link between pairs of nodes (A and B, say) whenever there is at least one link from a state of type A to a state of type B in the cell state network (see Fig. 1 for an example). An elementary result of graph theory states that the resulting network, which is known as the *condensation*, does not contain cycles (directed paths from a node to itself) and is, therefore, hierarchical (20). More precisely, we can order the cell types $T_1, T_2, \ldots$, such that if there is a trajectory (i.e. a directed path) from T_k to T_l, then $k \leq l$ (Fig. 1). In this case, we will say that T_k is *upstream* of T_l and T_l is *downstream* of T_k. Biologically, this means that cell types are necessarily ordered in a hierarchy, as is commonly observed.

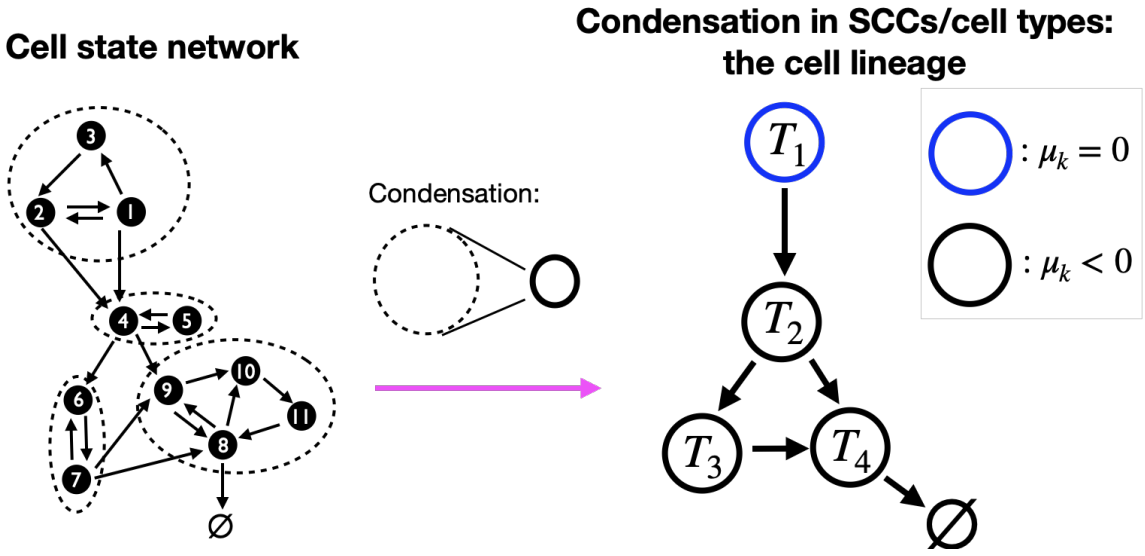

Fig. 1. Grouping of cell states into cell types. (Left) A hypothetical cell state network. In this network cell (molecular) states are represented as nodes and possible transitions between states as links. Cell loss (via death or emigration) is represented by the empty set symbol $\emptyset$. The dashed circles comprise states which are mutually reachable by directed paths, and thus are strongly connected components (SCCs) of the network. In our formulation, the SCCs represent cell types. (Right) When cell states are grouped into SCCs associated with cell types, the corresponding condensation forms a directed network without cycles, that encodes possible transitions between cell types. This network has a natural hierarchical structure: it admits an ordering of the nodes (cell types) $T_1, T_2, \ldots$ such that all transitions respect the ordering, that is, if there is a link, or a trajectory, from T_k to T_l then $k \leq l$. Here, we show one such hierarchical ordering (T_1 to T_4). Additionally, we show a possible distribution of transient (black circles) and self-renewing (blue circles) cell types T_k, $k = 1, \ldots, 4$, depending on their growth parameter μ_k (see Box 2). Finally, note that different choices for the discretization of the cell state trajectories will result in a different cell state network (left) but the same condensation network of cell types (right) —see main text.

B. Homeostasis constrains possible cell lineage hierarchies. So far we have considered the structure of the cell state and cell type networks without imposing any restriction on tissue proliferative dynamics. In the following, we consider tissues at homeostasis. Importantly, cell proliferation and removal must be finely balanced in renewing tissues to ensure homeostasis, and so homeostasis imposes strong constraints on proliferative dynamics. From a mathematical perspective, homeostasis represents a steady state of the cell population dynamics in which the number of cells of each type in the tissue stays, on average, constant over time. Our guiding question is: how does the dynamical constraint of homeostasis restrict the possible hierarchy of cell lineages?

B.1. Cell population dynamics. Since we are interested in cell state trajectories and cell population dynamics broadly defined, we generally need to consider three cell processes: cell division, direct transitions between cell states (i.e. changes in the internal molecular state of a cell), and cell loss (e.g. by cell death or emigration). Consider a cell in possible states $i = 1, 2, ..., m$, (discretized as discussed in the previous section). We can represent these three cell processes schematically as

$$\text{cell division: } i \to j + k \qquad [1]$$

$$\text{direct state transition: } i \to j \qquad [2]$$

$$\text{cell loss: } i \to \emptyset\,, \qquad [3]$$

where the empty set symbol $\emptyset$ indicates 'cell loss' and i, j, k denote cell states of which any two or all may be equal or not. The processes Eq. (1)–Eq. (3) determine the general dynamics of cells in a tissue. In this description, transitions between cell states can occur through cell divisions coupled to cell state changes, Eq. (1) (e.g. asymmetric cell division), and direct transitions, Eq. (2). Both transition types constitute the links of the cell state network. Equations describing the dynamics of the expected number of cells in each state can be easily derived from Eq. (1)–Eq. (3), and are provided in Box 1.

Box 1: Dynamics of expected cell numbers

Let $n_i(t)$ denote the expected number of cells in state i at time t. Furthermore, let us write the rate of cell division in state i as λ_i, the rate of direct transition from state i to state j as $\omega_{i\to j}$, the rate of cell loss in state i as d_i, and the probability that a cell in state i will produce daughter cells in state j and k when it divides as r_i^{jk}. The time evolution of

the expected number of cells of type i is given by,

$$\frac{d}{dt}n_i = \sum_j \left(\lambda_j r_j^i + w_{j\to i}\right) n_j - (\lambda_i + \sum_j w_{i\to j} + d_i)n_i, \qquad [4]$$

where $r_i^j = \sum_k (r_i^{jk} + r_i^{kj})$ is the expected number of cells in state j produced upon division of a cell in state i. We can write this system compactly in matrix form as follows. First, define the *total transition rate* from i to j as $\kappa_{i\to j} = \lambda_i r_i^j + \omega_{i\to j}$, which combines all transitions from state i to state j, via cell division, Eq. (1), and direct state transitions, Eq. (2). Note that $\kappa_{i\to j} \geq 0$ for all i, j. Also, define the *effective loss rate at* i as $\delta_i = \lambda_i + \sum_j \omega_{i\to j} + d_i$, which combines transitions out of state i and cell loss. Using this notation, we can express the dynamical system, Eq. (4), in matrix form as

$$\frac{d}{dt}\boldsymbol{n}(t) = A\,\boldsymbol{n}(t) \qquad [5]$$

where A is the $m \times m$ matrix with (i, j)-entry $\kappa_{i\to j}$ if $i \neq j$ and $\kappa_{i\to i} - \delta_i$ if $i = j$, and $\boldsymbol{n}(t) = (n_1(t), n_2(t), ..., n_m(t))$.

The matrix A encodes the weighted cell state network described in the main text as its adjacency matrix. This network contains all the information about the cell dynamics: the nodes represent the cell states $i = 1, 2, ..., m$ and any link from node (i.e. cell state) i to node j has weight $\kappa_{i\to j}$, representing how frequent this transition is. In particular, a link from cell state i to cell state j only exists if the transition is possible, that is, if $\kappa_{i\to j} > 0$. Note that self-loops (links from a node to itself) may occur, and that cell loss $\emptyset$ is not a node on the network, although we will add it to the graphical representation of our networks for clarity.

In general, the parameters λ_i, r_i^j, $\omega_{i\to j}$, d_i $(i, j = 1, ..., m)$ may depend on $n_1, n_2, \ldots$. If so, Eq. (4) will be non-linear. This case is discussed further below, with details provided in Box 3 (a complete mathematical treatment is given in the Supplemental Material, section 2). However, when the tissue is at homeostasis, the expected cell numbers do not change under the dynamics (that is, mathematically, $\frac{d}{dt}\boldsymbol{n} = 0$). In this case, the dynamics in both the linear and nonlinear regimes are determined by the properties of the matrix A, as we show in the Supplemental Material, section 1.

B.2. Dynamic classification of cell types. In order to assess conditions for tissue homeostasis we classify cell types according to their *intrinsic* long-term proliferative properties, that is, their intrinsic ability to perpetuate their own numbers, neglecting influx from other cell types. In Box 2, we show that the intrinsic long-term dynamics of the kth cell type population, T_k, can be uniquely characterised by a single number, its growth parameter μ_k, which is fully defined by the rates and probabilities of the processes depicted in Eq. (1)-Eq. (3). In particular, if $\mu_k > 0$ then the expected number of cells of type T_k increases; if $\mu_k < 0$ then the expected number of cells of type T_k decreases until the pool vanishes; and if $\mu_k = 0$ then the expected number of cells of type T_k stays constant, i.e. cells of that type maintain themselves at homeostasis (21). In general, μ_k may depend on the cellular environment and change over time (see discussion in section 2C), but in this section we refer to its steady state value only. This classification of cell types according to their intrinsic dynamics motivates the following definition:

Definition 2. *We call a cell type T with growth parameter μ:*

1. transient *if* $\mu < 0$,
2. hyper-proliferating *if* $\mu > 0$,
3. self-renewing *if* $\mu = 0$.

These three classes represent three distinctly different kinds of dynamics. Firstly, transient cell types will eventually vanish, unless they are replenished through incoming trajectories from other upstream cell types. This means that all transient cells and their progeny will eventually differentiate into another type or will be lost from the tissue (e.g. via cell death or emigration). Secondly, hyper-proliferating cell types will continually increase in number in the tissue. Although possible, this situation is clearly not physiological, yet may be encountered in certain pathologies, such as cancer. Thirdly, self-renewing cell types will maintain their numbers, on average, over time. In principle, this definition allows for inert cells that neither divide, differentiate or die. However, this situation only occurs in cell lineages that are not renewing, which we do not consider here. In renewing cell lineages, such cells must precisely balance proliferation and loss, for example by asymmetric cell divisions or neutral competition of equipotent cells for limited niche space (22–24), and so exhibit the biological property of self-renewal that is characteristic of stem cells. Such a fine-tuning at a critical point (an exact balance between differentiation and self-renewal) may seem at first glance implausible. However, in section 2C we give an example of a feedback-based regulation mechanism that can achieve such precise balance without intrinsic fine-tuning.

Box 2: Cell type dynamics

Consider a cell type T_k consisting of states $i_1, i_2, \ldots, i_{m_k}$ in the cell state network. The intrinsic dynamics of the cell type T_k, i.e. when neglecting transitions from other cell types, are determined by the equation

$$\frac{d}{dt}\boldsymbol{n}_k(t) = A_k \boldsymbol{n}_k(t), \qquad [6]$$

where $\boldsymbol{n}_k(t) = (n_{i_1}(t), n_{i_2}(t), ..., n_{i_{m_k}}(t))$ is the vector of expected cell numbers at time t in the cell states $i_1, i_2, \ldots, i_{m_k}$ comprising the cell type T_k. The matrix A_k is the $m_k \times m_k$ submatrix of A (see Box 1) corresponding to the rows and columns $i_1, i_2, ..., i_{m_k}$. Since A_k is the adjacency matrix of a strongly connected network (the SCC corresponding to T_k), it is irreducible, and furthermore, all off-diagonal elements of A are non-negative. Therefore, the Perron-Frobenius theorem guarantees that A_k has a simple real maximal eigenvalue μ_k (25, 26). It is a basic result of dynamical systems theory that μ_k determines the asymptotic long-term dynamics of the system, that is, $\frac{dn_k}{dt} \sim \mu_k n_k$ asymptotically, where n_k is the total number of cells in type T_k. This means that the sign of μ_k determines if the population will grow or decline in the long term (21) (see Section 1 of the Supplemental Material for further details). We call μ_k the *growth parameter* of cell type T_k.

B.3. Homeostasis imposes fundamental constraints on cell lineage architectures. There are a number of ways that homeostasis can be maintained. In principle, a homogeneous cell population that consists of a self-renewing population of a single type, as defined above, would, by definition, be homeostatic. However, this is a particular situation that is not reflective of the complexity of most tissues. Typically, a tissue will consist of a range of different cells, each with different proliferative capacities, related in a complex hierarchy. In these circumstances, the number of cells in each type are affected by incoming transitions from upstream cell types. Thus, for the cell population as a whole to be homeostatic, further conditions on the architecture of the lineage hierarchy need to be satisfied, which are discussed next.

The cell state dynamics described in Box 1 have the particular feature that the coupling between different states is non-negative, since transition rates $\kappa_{i\to j} \geq 0$. Furthermore, when considering the conditions for homeostasis – a steady state of the dynamics – the kinetic parameters can be treated as constants. Such a system is a *linear cooperative system*, i.e. a system where all couplings between distinct states are non-negative and constant. We have recently derived general mathematical conditions for a steady state in linear cooperative systems (27), and in the Supplemental Material (section 1) we show how this can be generalised to the non-linear dynamics described in Box 1. In particular, we show that homeostasis can only prevail in a renewing tissue if:

(i) No hyper-proliferating cell type is present.

(ii) At least one self-renewing cell type is present.

(iii) There are no cell state trajectories from one self-renewing cell type to another.

For now, we refrain from addressing the question how these conditions are maintained in order to achieve homeostasis. We will discuss this later, in section 2C, which outlines a mechanism that controls homeostasis.

Conditions (i-iii) are argued in a mathematically rigorous way in the Supplemental Material (section 1), yet they can also be understood intuitively. Condition (i) is obvious, since the presence of hyper-proliferating cells necessarily contradicts homeostasis. Condition (ii) follows from the fact that, in the absence of both hyper-proliferating and self-renewing cell types, all cell types must be transient and thus would steadily decline in numbers. Condition (iii), however, is less intuitive. To understand why a lineage connection between two self-renewing cell types disrupts homeostasis, let us consider a self-renewing type S_1 which is upstream of another one, S_2. If so, cells of type S_1 must be able to transition to type S_2. Given that the mean cell number of S_2 cells is constant without contribution from other cell types (by definition of a self-renewing type), this additional steady inflow of cells into type S_2 means that the number of S_2 cells must steadily increase, and thus be non-homeostatic. Therefore, self-renewing cell types cannot be connected by cell state trajectories, in a homeostasic state.

From conditions (i) and (iii) it follows that all cell types upstream of a self-renewing type must be transient. Therefore, they vanish in the long term from the tissue: that is, their cell numbers will tend to zero (see Ref. (27) and Supplemental Material, section 1). Since at least one self-renewing cell type must exist (condition (ii)), this implies – using our definition of a self-renewing cell type – that,

Principle 1. *In homeostatic renewing tissues, every self-renewing cell type resides at an apex of a cell lineage hierarchy, and every such lineage has a self-renewing type at its apex.*

Since self-renewing cell types reside at the apexes of lineage hierarchies, they can generate all cells of their lineage. Thus, self-renewal and lineage potential are not independent cellular properties of cells: they are fundamentally coupled. In particular, self-renewal implies lineage potential and lineage potential implies self-renewal. According to the commonly used characterisation of stem cells, we can therefore associate any self-renewing cell type in a homeostatic renewing tissue (with a non-trivial lineage) as a homeostatic *adult stem cell*. Moreover, stem cell identity can be established by either property.

The implications of conditions (i-iii) and Principle 1 are illustrated in Fig. 2, which shows a typical cell lineage hierarchy. Two features are notable.

First, Principle 1 allows cell types upstream of a self-renewing type to be transiently part of a lineage hierarchy in non-homeostatic situations – for example, during development or when a tissue is damaged – but they vanish from the tissue as homeostasis is approached or become dormant. Such *dormant* stem cells can reside above other self-renewing cells in non-homeostatic conditions, but they cannot participate in tissue renewal under homeostatic conditions and are therefore not considered part of the homeostatic lineage (see Fig. 2). Thus, our framework makes a clear distinction between stem cells that maintain homeostasis and those that regulate tissue response to conditions of stress.

Second, multiple apexes of a lineage can in principle exist, with different stem cells at each apex, as is conjectured for mouse mammary epithelium lineage, for instance (28).

With these insights, we can classify each cell type according to its growth parameter and relative position in the hierarchy into one of the following three classes.

Developing cells (D): These are transient cells (growth parameter $\mu < 0$) upstream of a self-renewing cell type in the lineage hierarchy. Developing cells may transiently divide and differentiate but, since they are not replenished by a self-renewing cell type, they disappear as homeostasis is approached and are thus not present in adult, homeostatic tissues. We therefore associate these cells with developmental populations that do not survive into adulthood (e.g. embryonic stem cells).

Committed cells (C): These are transient cells (growth parameter $\mu < 0$) downstream of a self-renewing cell type. They may divide transiently, but all their progeny are eventually lost (e.g. to death or emigration). However, this population does not disappear since it is maintained at homeostasis by replenishment from an upstream self-renewing population. We associate committed cells with adult somatic cells that are not stem cells.

Adult stem cells (S): These are self-renewing cells (growth parameter $\mu = 0$, maintaining their population) which can differentiate into at least one committed cell type downstream. According to Principle 1, they must reside at the top of a homeostatic lineage hierarchy.

Principle 1 and the conditions (i-iii) above have a number of implications. For example, it immediately follows that only three patterns of stem cell division are allowed in homeostasis:

$$S_k \to S_k + S_k, \quad [7]$$

$$S_k \to S_k + C_l, \quad [8]$$

$$S_k \to C_l + C_m, \quad [9]$$

where S_k is an adult stem cell type of type k, and C_l, C_m are (possibly different) committed cell types. These three division types correspond precisely to what is known experimentally about stem cell divisions in homeostasis. In particular, stem cells may divide in the following three ways: (1) symmetric self-renewal divisions ($S \to S + S$), (2) asymmetric self-renewal divisions ($S \to S + C$) and (3) differentiation division ($S \to C + C$). However, following the conditions for homeostasis above, we can say more. For instance, a stem cell division producing one or more stem cells of different types, i.e. a division of the form $S_k \to S_l + S_m$ or $S_k \to S_l + C$ in which $l \neq k$ or $m \neq k$ is not possible, as it would contradict condition (iii) above. Furthermore, stem cell divisions that produce developing cells, and divisions of a committed cell to produce a stem cell, are also not possible, according to Principle 1. To make these ideas more concrete, examples of lineage hierarchies that do/do not fulfil these principles are shown in Figures 2 and 3.

Collectively, these results demonstrate that characteristic properties of lineage architectures may be derived from first principles via consideration of the relationship between cell states and cell types, and from dynamical constraints imposed on lineage hierarchies by homeostasis.

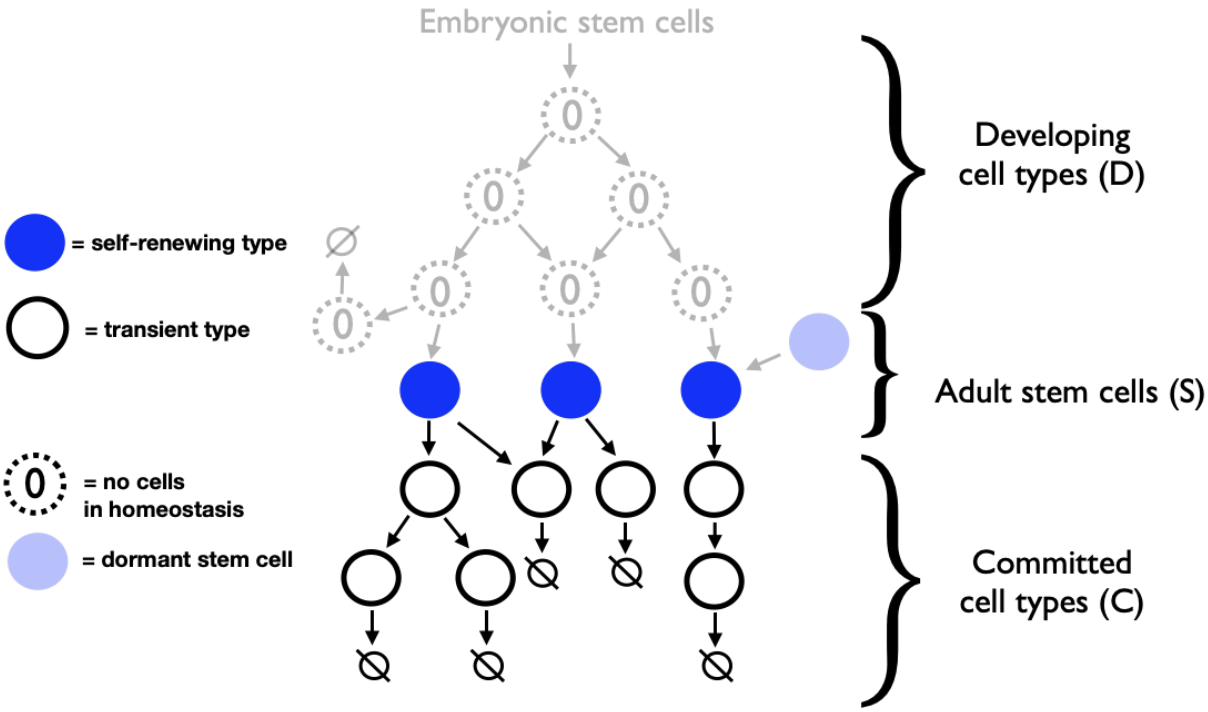

Fig. 2. Illustration of a typical cell lineage tree according to Principle 1. Each circle represents a cell type (see Fig. 1), which comprises a maximal set of mutually reachable cell states, and arrows are possible transitions between cell types. Filled-in (blue) circles represent self-renewing cell types. Black bold circles are committed cell types (C) which can differentiate and whose progeny is eventually lost. Cell types contributing to homeostasis are in full colours. Faint circles and arrows are cell types and trajectories which do not contribute to homeostatic tissue renewal: Faint grey dashed circles are developing cell types (D) which are present during development, but which disappear in the adult tissue. Faint filled-in blue circles are dormant stem cells which can become activated while the tissue is disturbed (non-homeostatic), but are not active during homeostasis. Crucially, along each homeostatic lineage trajectory (a series of transitions between cell types that are active in homeostasis) only a single self-renewing cell type can contribute to homeostasis, which we identify as adult stem cells (S). Thus, in the long term, a single stem cell type must be at each apex of the homeostatic lineage.

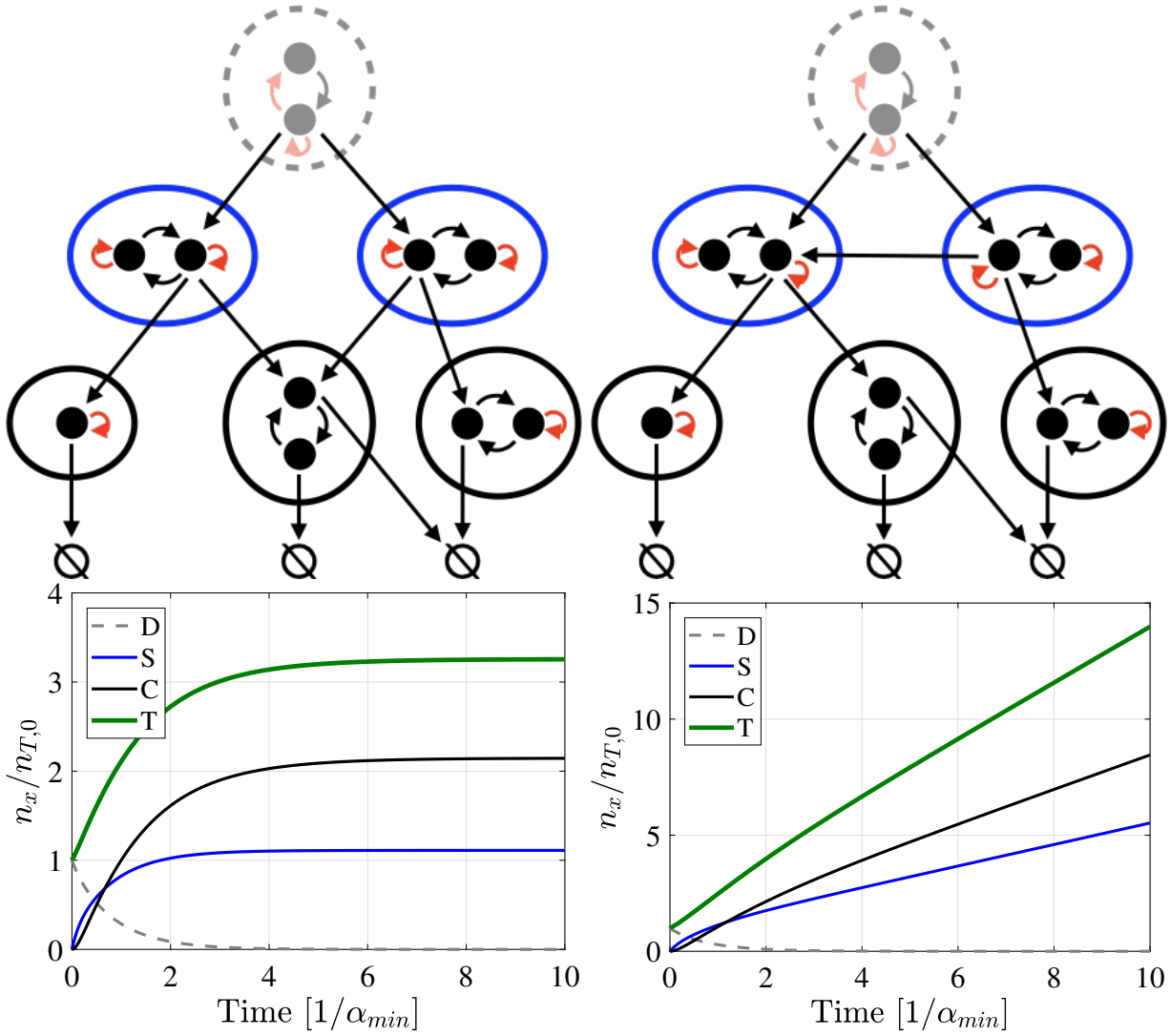

Fig. 3. Lineage hierarchy and cell dynamics. (Top) Two cell state networks: black dots represent cell states, black arrows represent direct cell state transitions (according to Eq. (2)), red arrows represent transitions through cell divisions (according to Eq. (1)), and $\emptyset$ denotes cell loss (via cell death or emigration, for example). We draw circles around cell states of the same type (the SCCs of the underlying cell state network): black circles represent transient cell types, and blue circles self-renewing cell types. Grey faint states and circles denote developing cell states and types. (Bottom) Simulations of expected cell numbers obtained by solving Eq. (4) numerically for the networks immediately above ($\alpha_{min} = \min_{i,j}(\lambda_i, \omega_{i\to j}, d_i)$ is the inverse of the smallest rate parameter). (Left panels) This network (top left) fulfils the conditions (i-iii) for homeostasis, and thus (bottom left) the expected numbers of stem cells (n_S), and committed cells (n_C) reach a steady state, while the expected number of developing cells (n_D) tends to zero. The total expected number of cells, n_T, also reaches a steady state. (Right panels) This network (top right) does not fulfil the condition (iii) for homeostasis since two self-renewing cell types are connected by a direct link and thus (bottom right) the cell numbers n_S and n_C diverge.

C. Environmental regulation of the stem cell identity. In order to be self-renewing, the kinetic parameters of a stem cell type (rates of division, cell state transitions, and cell loss) need to be finely tuned to achieve a growth parameter of exactly zero, ($\mu = 0$, see Box 2) and to ensure homeostasis. Any, even slight, deviation from this value may, if sustained over an extended period, lead to a loss of the self-renewing capacity and thus of the stem cell phenotype. In the absence of a homeostatic control mechanism such fine tuning is biologically implausible. In reality, cells are embedded in a tissue environment and may respond to signalling cues from other cells. A simple example of such regulation is crowding feedback, in which cells sense the density of cells in their local micro-environment and respond by adjusting their propensity to divide or differentiate. Such feedback mechanisms have been experimentally observed in various epithelia such as in developing *Drosophila* (29), human colon and zebrafish epidermis (30, 31), where increased cell density is known to accelerate cell differentiation. Potential mechanisms to mediated this feedback are by mechanosensing (31–33) or by the limited availability of growth signalling factors (24), for example.

We can incorporate such regulation in our framework by assuming that the kinetic parameters associated with the dynamics in Eq. (1)-Eq. (3), as defined in Box 1, are not constant, but depend on the cell density. A kinetic parameter may increase with the cell density (i.e. exhibit *positive crowding dependence*) or decrease with it (exhibit *negative crowding dependence*), or neither. In Box 3 below, and in section 2 of the Supplemental Material, we show that a simple sign criterion for the crowding dependence of the kinetic parameters guarantees that the cell type at the apex of the cell lineage hierarchy will acquire and maintain self-renewal potential, on average, in the long term. In short, the sign criterion ensures that the growth parameter μ decreases with cell density, and thus with cell number N (i.e. $d\mu/dN < 0$).

Recall that the growth parameter μ determines the long term growth rate of the cell population (i.e. asymptotically, $\mu \sim dN/dt$). Therefore, the sign criterion ensures that if the cell number is below the critical value N^* for which $\mu = 0$, then $\mu > 0$ and thus the cell number eventually increases toward N^*. Similarly, if the cell number is above N^*, then $\mu < 0$, and thus the cell number eventually decreases toward N^*. Hence, the tissue cell numbers will be confined around the self-renewing state defined by $\mu = 0$ and are either stable, or fluctuate or oscillate around the homeostatic state (for more details, see the discussion in the Supplemental Material, section 2). Notably, this could include a dynamic state, in which cells persistently switch between a hyper-proliferating state ($\mu > 0$) and a declining state ($\mu < 0$), in a way, however, that *on average* cell numbers remain in the long term constant. Either way, the cell population is on average long-term self-renewing, resulting in a (potentially *dynamic*) homeostatic state for the tissue, assuming all other cell types are transient.

This feedback-based mechanism for regulating stemness is simple and robust against noise, as it does not depend on the exact values of the dynamic parameters, but only on the sign of their dependencies. Crucially, this means that self-renewal need not be a property intrinsically determined by a cell, but it may be acquired by any cell type that is at the apex of a lineage hierarchy, through interaction with the cellular environment.

Box 3: Regulation of self-renewal by crowding feedback

Consider a cell type T_a at the apex of a cell lineage, which is not a priori a self-renewing type. Assume that the kinetic parameters of the cell dynamics, λ_i, $\omega_{i\to j}$, and d_i (see Box 1), may depend on the density of cells in the 'niche', which we assume to be constituted by the cells of type T_a. For constant volume, the density is proportional to the number of cells of type T_a, $N_a = \sum_{i\in T_a} n_i$, and thus these parameters are functions of N_a: $\lambda_i(N_a)$, $\omega_{i\to j}(N_a)$, and $d_i(N_a)$. The time evolution for cell numbers per state in T_a, Eq. (6), then becomes non-linear: $\frac{d}{dt}\boldsymbol{n}_a = A_a(N_a)\boldsymbol{n}_a$. We note that, since the kinetic parameters are functions of N_a, so are the entries of the matrix A_a and thus the growth parameter $\mu_a = \mu_a(N_a)$ as well.

Now, for notational convenience, let us rename all kinetic parameters by α_j, numbered by $j = 1, ..., m_a^2 + m_a$, in the order $\lambda_1, \lambda_2, ..., \omega_{1\to 2}, \omega_{1\to 3}, ..., d_1, d_2, ...$ In Section 2 of the Supplemental Material, we show that a sufficient condition for the number of cells of type T_a to neither increase nor decrease monotonically in the long term (i.e. to obtain either a *stable* self-renewing state or exhibit finite oscillations/fluctuations around it) is

$$\frac{d\mu_a}{dN_a} = \sum_j \frac{d\mu_a}{d\alpha_j}\frac{d\alpha_j}{dN_a} < 0\,, \qquad [10]$$

for all N_a (Furthermore, we make the biologically plausible assumption that for large N_a ($N_a \to \infty$) each parameter attains a limiting value). Recall that the crowding dependence of a parameter α is positive if $d\alpha/dN_a > 0$ and negative if $d\alpha/dN_a < 0$. For a cell type at the apex of a non-trivial lineage hierarchy the following sufficient criterion holds:

Sign criterion for regulation of self-renewal: If for all kinetic parameters α_j, the crowding dependence is positive (respectively negative), while the growth parameter dependence on α_j is negative (respectively positive), then a cell type at the apex of a lineage hierarchy acquires and maintains a long-term self-renewing state (i.e. the state of self-renewal is stable, or cell numbers fluctuate/oscillate around that state).

Note that the sign criterion is sufficient but not necessary, and not all parameters must fulfil the sign criterion. Eq. (10) can also be satisfied if some parameters do not meet this requirement, as long as they do not outweigh those which do. In particular, for any parameter with $d\mu_a/d\alpha_j = 0$, the crowding dependence of α_j is not relevant. Illustrative examples of this result are shown in the Supplemental Material, Figure 1.

3. Discussion

Stem cells are fundamental to healthy tissue turnover and repair and have accordingly been the subject of concerted research. However, despite more than 60 years of study, our understanding of the general principles that underpin stem cell dynamics remains incomplete. Here, we have proposed a new way to approach understanding stem cell identities from first principles based upon consideration of the relationships between cell states and cell types and general properties of proliferation in hierarchically structured populations. The resulting organisational principles are consistent with common conceptualisations of adult stem cells and lineage architectures, even though they emerge from consideration of the dynamical properties of cell lineage architectures at homeostasis rather than from explicitly biological considerations.

Furthermore, our analysis indicates that these common properties are not coincidental but rather reflect fundamental and universal constraints on tissue dynamics at homeostasis – defined here as a steady state of the tissue cell population. The common observation that self-renewing cells sit at the apex of lineage hierarchies, for example, is not simply due to biological contingency: it is the only possible way that lineage architectures can be constructed to support homeostasis in renewing tissues. Crucially, this means that lineage potential and self-renewal potential are not independent characteristics of adult stem cells in homeostasis, but they imply each other. Thus, in homeostasis, either self-renewal or lineage potential alone is enough to define a stem cell.

It is important to note that our results only hold at homeostasis. In challenged conditions, such as wounding and cancer, lineage architectures may temporarily change to allow de-differentiation, for example. In such situations properties of the lineage hierarchy not accounted for in our formulation may also be important. For example, our Principle 1 does not exclude an additional pool of dormant stem cells residing in the lineage hierarchy above an adult stem cell population, such as a dormant population of epidermal stem cells residing in hair follicles (34). However, our analysis shows that such dormant stem cells can not participate in homeostatic tissue renewal if another self-renewing cell type is present, and so they play no role in healthy adult tissue turnover. They may nevertheless become activated under conditions of stress.

Our results accord well with current experimental evidence in numerous tissues. For example, in many mammalian renewing epithelia, such as mouse inter-follicular epidermis (22, 35), oesophagus (36, 37), and intestine (38), a single self-renewing cell type, at the apex of the homeostatic lineage hierarchy, can be identified. In some tissues, however, multiple stem cell populations that contribute to the same pool of cells have been identified, such as in the hematopoietic lineage (39), human epidermis (34), sebaceous glands (40), or as conjectured for the mammary gland (28). While not immediately apparent, these cases are generally consistent with our predictions. In the mouse mammary gland, for example, the two stem cell populations each reside on a separate apex of the lineage hierarchy (28), and thus are allowed in our framework. In other cases, additional stem cell populations represent dormant stem cells (as discussed above for epidermis (34)), or stem cells that are most likely active only in development or under conditions of stress (e.g. Blimp1+ sebaceous stem cells which are distinct from self-renewing progenitor cells (41), and which can recreate sebaceous organoids (40)), or stem cells that trans-differentiate between compartments (in epidermis (42)). In each of these examples the activity of multiple stem cells only occurs in a non-homeostatic context – a situation not covered by our theory. An intriguing potential counter-example is the proposition of two distinct types of stem cells in the homeostatic hematopoietic lineage – long-term (LT) and short-term (ST) repopulating HSCs – since it has been suggested that they stand in a hierarchical relationship to one another (39). However, much of our understanding of HSC function comes from bone marrow transplantation experiments, which are fundamentally out of homeostasis, and the functional distinction between LT-HSCs and ST-HSCs has accordingly not yet been fully characterised *in vivo*. Indeed, there is increasing evidence that the mechanism of HSC proliferation in times of stress is fundamentally different to that under steady state. Our analysis could be a guide to future experiments since Principle 1 predicts that either: (1) some identified HSCs constitute a dormant population that does not participate in normal homeostasis, but are activated when homeostasis is

disturbed, e.g. after blood loss; or (2) one of the putative stem cell populations is not genuinely self-renewing, but gradually loses its proliferative potential on a long time scale. (3) LT-HSCs and ST-HSCs can inter-convert (even if only rarely). If so, they would, according to our definition, effectively be a single stem cell type. These issues are hard to resolve experimentally, yet theoretical considerations can be a guide. It is likely that similar issues will arise as we gain deeper understanding of other systems. If so, then theoretical notions such as the ones we propose could be invaluable. We anticipate that future developments in stem cell biology will benefit from closer integration of experiment and theory which will yield a deeper understanding of the organisational principles that regulate stem cell dynamics.

Materials and Methods

The illustrative cell dynamics shown in Figure 3 and in Figure 1 of the Supplemental Material are obtained by numerical integration based on Runge-Kutta 4-5 method of the system of ODEs for the corresponding networks. The code is implemented in Matlab and available at https://github.com/cp4u17/stemCellIdentity.git.

ACKNOWLEDGMENTS. P.G. is supported by a Medical Research Council New Investigator Research Grant MR/R026610/1, C.P. is supported by a Studentship of the Institute for Life Sciences (Southampton), and P.G. and R.S.-G. are partially supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

1. Siminovitch L, McCulloch EA, Till JE (1963) The Distribution of Colony-Forming Cells Among Spleen Colonies. *Journal of cellular physiology* 62(3):327.
2. Becker AJ, McCulloch EA, Till JE (1963) Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* 197(4866):452–454.
3. Blau HM, Brazelton TR, Weimann JM (2001) The evolving concept of a stem cell: Entity or function? *Cell* 105(7):829–841.
4. Seaberg RM, Van Der Kooy D (2003) Stem and progenitor cells: The premature desertion of rigorous definitions. *Trends in Neurosciences* 26(3):125–131.
5. Shostak S (2006) (Re)defining stem cells. *BioEssays* 28(3):301–308.
6. Clevers H (2015) What is an adult stem cell? 350(6266):4–6.
7. Laplane L (2015) Stem cell epistemological issues. *Stem Cell Biology and Regenerative Medicine* (Umr 1009):693–712.
8. Vorotelyak E, Vasiliev A, Terskikh V (2020) The Problem of Stem Cell Definition. *Cell and Tissue Biology* 14(3):169–177.
9. Ghadially R (2012) 25 years of epidermal stem cell research. *Journal of Investigative Dermatology* 132:797.
10. Alcolea MP (2017) *Oesophageal Stem Cells and Cancer*, ed. Birbrair A. (Springer International Publishing, Cham), pp. 187–206.
11. Donati G, Watt FM (2015) Stem Cell Heterogeneity and Plasticity in Epithelia. *Cell Stem Cell* 16(5):465–476.
12. Donati G, et al. (2017) Wounding induces dedifferentiation of epidermal Gata6+ cells and acquisition of stem cell properties. *Nature Cell Biology* 19(6):603–613.
13. Tetteh PW, Farin HF, Clevers H (2015) Plasticity within stem cell hierarchies in mammalian epithelia. *Trends in Cell Biology* 25(2):100–108.
14. Tetteh PW, et al. (2016) Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* 18(2):203–213.
15. Koren S, Bentires-Alj M (2015) Review Breast Tumor Heterogeneity : Source of Fitness , Hurdle for Therapy. *Molecular Cell* 60(4):537–546.
16. Van Keymeulen A, et al. (2015) Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* 525:119.
17. Clevers H, Watt FM (2018) Defining adult stem cells by function, not by phenotype. *Annual review of biochemistry* 87:1015.
18. Milnor J (2016) *Morse theory.(AM-51).* (Princeton university press) Vol. 51.
19. Bollobás B (2013) *Modern graph theory.* (Springer Science & Business Media) Vol. 184.
20. Cormen TH (2009) *Introduction to Algorithms.* (MIT Press).
21. Strogatz S (1994) *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, engineering.* (Perseus Books) Vol. 32, pp. 32–0994–32–0994.
22. Clayton E, et al. (2007) A single type of progenitor cell maintains normal epidermis. *Nature* 446:185.
23. Lopez-Garcia C, Klein AM, Simons BD, Winton DJ (2010) Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. *Science* 330:822.
24. Kitadate Y, et al. (2019) Competition for Mitogens Regulates Spermatogenic Stem Cell Homeostasis in an Open Niche. *Cell Stem Cell* 24(1):79–92.
25. MacCluer CR (2000) The Many Proofs and Applications of Perron's Theorem. *SIAM Review* 42(3):487–498.
26. Arrow KJ (1989) A "dynamic" proof of the Frobenuis-Perron theorem for Metzler matrices. *Probability, Statistics, and Mathematics* (6):17–26.
27. Greulich P, MacArthur BD, Parigini C, Sánchez García RJ (2019) Stability and steady state of complex cooperative systems: a diakoptic approach. *Royal Society Open Science* 6:191090.
28. Pal B, et al. (2017) Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nature Communications* 8(1):1627.
29. Marinari E, et al. (2012) Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature* 484:542.
30. Eisenhoffer GT, et al. (2012) Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia. *Nature* 484(7395):546–549.
31. Eisenhoffer GT, Rosenblatt J (2013) Bringing balance by force: live cell extrusion controls epithelial cell numbers. *Trends in cell biology* 23(4):185–92.
32. Puliafito A, et al. (2012) Collective and single cell behavior in epithelial contact inhibition. *Proc. Natl. Acad. Sci.* 109:739.
33. Shraiman BI (2005) Mechanical feedback as a possible regulator of tissue growth. *Proc. Natl. Acad. Sci.* 102:3318.
34. Taylor G, Lehrer MS, Jensen PJ, Sun TT, Lavker RM (2000) Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* 102(4):451–461.
35. Alonso L, Fuchs E (2003) Stem cells of the skin epithelium. *Proc. Natl. Acad. Sci.* 100:11830.
36. Seery JP, Watt FM (2000) Asymmetric stem-cell divisions define the architecture of human oesophageal epithelium. *Current Biology* 10(22):1447–1450.
37. Seery JP (2002) Stem cells of the oesophageal epithelium. *Journal of Cell Science* 115(9):1783–1789.
38. Barker N, et al. (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449(7165):1003–1007.
39. Cheng H, Zheng Z, Cheng T (2019) New paradigms on hematopoietic stem cell differentiation. *Protein and Cell* 11:34.
40. Feldman A, et al. (2019) Blimp1+ cells generate functional mouse sebaceous gland organoids in vitro. *Nature Communications* 10(1):1–13.
41. Andersen MS, et al. (2019) Tracing the cellular dynamics of sebaceous gland development in normal and perturbed states. *Nature Cell Biology* 21(8):924–932.
42. Jones PH, Simons BD, Watt FM (2007) Sic transit gloria: farewell to the epidermal transit amplifying cell? *Cell stem cell* 1(4):371–81.

Universal principles of lineage architecture and stem cell identity in renewing tissues – Supplemental Material

Philip Greulich[1,2], Ben D. MacArthur[1,3], Cristina Parigini[1,2], and Rubén J. Sánchez-García[1,2]

[1]Mathematical Sciences, University of Southampton, Southampton, United Kingdom
[2]Institute for Life Sciences, University of Southampton, Southampton, United Kingdom
[3]Centre for Human Development, Stem Cells and Regeneration, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

March 18, 2021

1 Conditions for homeostasis on lineage hierarchies

In the following we derive the conditions for homeostatic tissue cell dynamics by adapting conditions for the existence of a steady state in linear cooperative systems (Ref. [1]) to the biological context. This section will deal only with conditions for a homeostatic state to prevail, without specifying whether it is stable. Section 2 of this document will then discuss a particular mechanism to render homeostasis stable.

First, we note that homeostasis is defined as a steady state of the corresponding expected cell numbers in each state, $\boldsymbol{n} = (n_1, n_2, ...)$. A steady state, $\boldsymbol{n}^*$, is a configuration of expected cell numbers that do not change over time, that is, $\frac{d\boldsymbol{n}}{dt} = 0$ for $\boldsymbol{n} = \boldsymbol{n}^*$. The tissue cell dynamics is described in the Main Text as Eq. **[5]** in Box 1. In general, this system may be non-linear, if its parameters depend on the state vector of cell numbers $\boldsymbol{n}$, therefore Eq. **[5]** in the Main Text may be written in full generality as, $\frac{d\boldsymbol{n}}{dt} = \tilde{A}(\boldsymbol{n})\boldsymbol{n}$, that is, the matrix $\tilde{A}$ depends on the dynamical variables (cell numbers) $n_1, n_2, ...$ Nonetheless, the condition for a cell number configuration, $\boldsymbol{n}^*$, to be a steady state – without specifying the stability of that steady state – reads, $\tilde{A}(\boldsymbol{n}^*)\,\boldsymbol{n}^* = 0$, which is identical to the steady state condition of the corresponding linear system $\frac{d\boldsymbol{n}}{dt} = A\,\boldsymbol{n}$ with $A = \tilde{A}(\boldsymbol{n}^*)$ being constant. Thus, the conditions for $\boldsymbol{n}^*$ to be a steady state are the same as that of the corresponding linear system $\frac{d\boldsymbol{n}}{dt} = A\,\boldsymbol{n}$, where $A = \tilde{A}(\boldsymbol{n}^*)$. The latter is a *linear cooperative system* (LCS) since all off-diagonal elements of A (and, in fact, of $\tilde{A}(\boldsymbol{n})$ for any $\boldsymbol{n}$) are non-negative (division and transition rates $\lambda_i, \omega_{ij}, d_i$ as defined in the Main Text, Box 3, are never negative) [2]. Hence, we can apply the conditions for the existence of a steady state of LCS, formulated and proven in Ref. [1], to the biological scenario of homeostasis. In order to use those conditions, let us introduce some mathematical definitions and terminology from Ref. [1].

- If A is an $n \times n$ matrix, $G(A)$ is the *graph* (or *network*) with adjacency matrix A, that is, the graph with n nodes and a link from node i to node j weighted by a_{ij}, the (i, j)-element of A. By definition, such a link only exists if $a_{ij} \neq 0$.

- We say that nodes i and j are *strongly connected* if there exists a directed path from node i to node j and from node j to node i. This is an equivalence relation on the set of nodes. The maximal subsets of strongly connected nodes (the equivalence classes of this equivalence relation) are called the *strongly connected components (SCC)* of the graph. Moreover, we identify an SCC with the subgraph generated by its nodes (the subgraph containing all the nodes, and all the links between them).

- The network $G(A)$ can be uniquely decomposed into its SCC (subgraphs) $S_1, S_2, ..., S_h$ [3]. Each connected component S_k is a graph with adjacency matrix B_k: namely, if S_k has n_k nodes, B_k is the $n_k \times n_k$ submatrix of A consisting of the rows and columns corresponding to the nodes of S_k. Each B_k has non-negative off-diagonal elements (because it is a submatrix of A), that is, it is a so-called *Metzler matrix*. Moreover, B_k is irreducible (since S_k is strongly connected). Therefore, the Perron-Frobenius theorem ensures that a unique, simple and real maximal eigenvalue μ_k exists [4]. This is called the *dominant eigenvalue* of B_k, and corresponds to the growth parameter defined in the Main Text. We call an SCC S_k *critical* if $\mu_k = 0$, *sub-critical* if $\mu_k < 0$, and *super-critical* if $\mu_k > 0$. For simplicity, we identify each SCC with its set of nodes, also written as S_k for $k = 1, 2, \ldots, h$.

- Given two SCC, S_k and S_l, if there is a (directed) path from a node in S_k to a node in S_l, we say that S_k is *upstream* of S_k and that S_l is *downstream* of S_k. (They are both partial orders in the set of SCC of $G(A)$.) Note that there cannot be (directed) paths from S_k to S_l *and* from S_l to S_k: otherwise, all nodes in both components would be mutually reachable and thus S_k and S_l would merge into a single SCC.

- We call a vector $\boldsymbol{x}$ *non-negative*, if $x_i \geqslant 0$ for all $i = 1, 2, ...$ We write this as $\boldsymbol{x} \geqslant 0$. Similarly a vector is *non-positive* if $x_i \leqslant 0$ for all $i = 1, 2, ...$ (written $\boldsymbol{x} \leqslant 0$), *positive* if $x_i > 0$ for all $i = 1, 2, ...$ (written $\boldsymbol{x} > 0$), and *negative* if $x_i < 0$ for all $i = 1, 2, ...$ (written $\boldsymbol{x} < 0$).

- A *steady state* of a dynamical system with dynamical variables $\boldsymbol{x} = (x_1, x_2, ...)$, is a vector $\boldsymbol{x} = \boldsymbol{x}^*$ for which $\frac{d}{dt}\boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}^*} = 0$. Furthermore, since we are only considering non-negative quantities (number of cells in each state), we require that $\boldsymbol{x}^*$ is non-negative. Clearly, the zero vector $\boldsymbol{x}^* = \boldsymbol{0}$ is a steady state, called the *trivial* steady state. A steady state is called (Lyapunov) *stable* if a small initial deviation from $\boldsymbol{x}^*$ at time $t = t_0$ leads to small deviations $\boldsymbol{x}(t)$ at any time $t > t_0$. More accurately: there exists a constant $C > 0$ such that $|\boldsymbol{x}(t) - \boldsymbol{x}^*| < C|\boldsymbol{x}_0 - \boldsymbol{x}^*|$, for all times $t \geqslant t_0$, where $\boldsymbol{x}_0 = \boldsymbol{x}(t = t_0)$ is the initial condition, sufficiently close to $\boldsymbol{x}^*$.

- An SCC S_k is called *vanishing*[1], if the steady state restricted to the nodes of that SCC is zero, i.e. $\boldsymbol{x}^*|_{S_k} = (x^*_{i_1}, x^*_{i_2}, ...)|_{i_1, i_2, ... \in S_k} = \boldsymbol{0}$. Otherwise, the SCC S_k is called *non-vanishing*. An SCC S_k is called *positive* if $\boldsymbol{x}^*|_{S_k}$ is a positive vector, that is, $x^*_i > 0$ for all $i \in S_k$.

According to Ref. [1], the following are necessary and sufficient conditions for the existence of a non-trivial, non-negative, (Lyapunov) stable steady state in a LCS.

Theorem 1. *The LCS $\frac{d}{dt}\boldsymbol{x} = A\boldsymbol{x}$ possesses a non-trivial, non-negative, Lyapunov stable, steady state if and only if*

1. *G(A) does not contain any super-critical SCC.*
2. *G(A) contains at least one critical SCC.*
3. *There are no directed paths from a critical SCC to another in $G(A)$. (Equivalently, there are no other critical SCC downstream, or upstream, of any critical SCC.)*

Furthermore, it can be shown that [1]:

Theorem 2. *Any SCC which is upstream of a critical SCC must be vanishing.*

For a general system $\frac{d\boldsymbol{x}}{dt} = A(\boldsymbol{x})\boldsymbol{x}$, we can show that the Conditions 1, 2 and 3 above are necessary for the existence of a steady state, as follows.

[1] We deviate at this point from the terminology in Ref. [1], where this property is called *trivial*, in order to avoid confusion with the same term used in the Main Text in a different context.

Theorem 3. *Consider a dynamical system of the form $\frac{d}{dt}\boldsymbol{x} = A(\boldsymbol{x})\boldsymbol{x}$, where $A(\boldsymbol{x})$ is a Metzler matrix (i.e. with non-negative off-diagonal elements) for all non-negative vectors $\boldsymbol{x} \geqslant 0$. If the system has a non-trivial, non-negative steady state $\boldsymbol{x}^*$ then the graph $G(A(\boldsymbol{x}^*))$ satisfies the following three conditions.*

1. *There are no non-vanishing super-critical SCC.*
2. *There is at least one non-vanishing critical SCC.*
3. *There are no directed paths from a non-vanishing critical SCC to another. (Equivalently, there are no other non-vanishing critical SCC upstream of any critical SCC.)*

Moreover, any SCC above a critical SCC must be vanishing, and every SCC is either vanishing, or positive.

We note that this theorem (to be proven below) does not consider the stability of the steady state, but only the existence of a non-trivial, non-negative, steady state, and it only refers to non-vanishing SCC, that is, SCC S_k which possess a non-zero contribution to the steady state, $\boldsymbol{x}^*|_{S_k} \neq \mathbf{0}$. In the biological context, however, a vanishing sub-system means that there are no cells of that type in the steady state, so a vanishing cell type – that is, the associated SCC in our model – is virtually non-existent in the real world context.

Proof of Theorem 3. Let us assume the existence of a non-trivial, non-negative steady state $\boldsymbol{x}^*$, that is, $\mathbf{0} \neq \boldsymbol{x}^* \geqslant 0$ and $A\boldsymbol{x}^* = \mathbf{0}$, where $A = A(\boldsymbol{x}^*)$. We assume a topological ordering of the nodes (see [1, Eq. 2.3]) so that the matrix A is a lower triangular matrix and we can write

$$A\boldsymbol{x} = \begin{pmatrix} B_1 & 0 & 0 & 0 & ... \\ C_{21} & B_2 & 0 & 0 & ... \\ C_{31} & C_{32} & B_3 & 0 & ... \\ \vdots & \vdots & \vdots & \ddots & 0 \\ ... & ... & ... & ... & B_h \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1^* \\ \boldsymbol{x}_2^* \\ \boldsymbol{x}_3^* \\ \vdots \\ \boldsymbol{x}_h^* \end{pmatrix} = \mathbf{0}, \tag{1}$$

where h is the number of SCC of $G(A)$, B_k is the adjacency matrix of the k-th SCC, S_k ($1 \leqslant k \leqslant h$), C_{kl} encodes the connectivity from S_l to S_k, and we have decomposed the steady state vector as $\boldsymbol{x}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, ..., \boldsymbol{x}_h^*)^T$. In particular, the kth row gives

$$\sum_{l<k} C_{kl}\boldsymbol{x}_l^* + B_k\boldsymbol{x}_k^* = \mathbf{0}. \tag{2}$$

For convenience, we write $\boldsymbol{y}_k = \sum_{l<k} C_{kl}\boldsymbol{x}_l^*$ and rewrite Eq.(2) as

$$\boldsymbol{y}_k + B_k\boldsymbol{x}_k^* = \mathbf{0}. \tag{3}$$

As explained before, each B_k is an irreducible Metzler matrix, so the Perron-Frobenius theorem applies [4], and thus B_k has a simple real eigenvalue μ_k of maximal real part among all eigenvalues of B_k. Moreover, μ_k has a positive (left) eigenvector $\boldsymbol{v}_k$, that is, $\boldsymbol{v}_k B_k = \mu_k \boldsymbol{v}_k$ and $\boldsymbol{v}_k > 0$. If we multiply Eq.(3) by the left eigenvector $\boldsymbol{v}_k$ we obtain

$$0 = \boldsymbol{v}_k\boldsymbol{y}_k + \boldsymbol{v}_k B_k\boldsymbol{x}_k^* \implies \boldsymbol{v}_k\boldsymbol{y}_k = -\mu_k\boldsymbol{v}_k\boldsymbol{x}_k^*. \tag{4}$$

In the following, we make repeatedly use of,

Lemma 1. *Let $\boldsymbol{z}, \boldsymbol{x} \in \mathbb{R}^n$. If $\boldsymbol{z} > 0$ and $\boldsymbol{x} \geqslant 0$, then holds:* $\boldsymbol{z}\boldsymbol{x} = 0 \Leftrightarrow \boldsymbol{x} = \mathbf{0}$.

This holds since $0 = \boldsymbol{z}\boldsymbol{x} = \sum_i z_i x_i$ requires that all $z_i x_i = 0$, given that all $z_i, x_i \geqslant 0$. Since $z_i > 0$, this can only be if $x_i = 0$ for all $i = 1, 2, ...$ The direction $\boldsymbol{x} = 0 \Rightarrow \boldsymbol{z}\boldsymbol{x} = 0$ of the lemma is trivial.

Now we distinguish the cases $\mu_k > 0, \mu_k = 0$, and $\mu_k < 0$.

Case $\mu_k > 0$. In this case (a super-critical SCC), $\mu_k > 0$, and $\boldsymbol{v}_k > 0$, $\boldsymbol{x}_k^* \geqslant 0$, so the right hand side of Eq. (4) is a non-positive number. Since $\boldsymbol{y}_k \geqslant 0$ and $\boldsymbol{v}_k > 0$, the left-hand side is a non-negative number. Hence both the left- and right-hand-side must be zero. Therefore, according to Lemma 1, $\boldsymbol{x}_k^*$ is zero, that is, the SCC is vanishing. Thus, *all super-critical SCC – if there are any – are vanishing*, which is **Condition 1**. Moreover, we get from Lemma 1 that $\boldsymbol{y}_k = \boldsymbol{0}$ for super-critical SCC, which we will use later.

Case $\mu_k < 0$. In this case (a sub-critical SCC), the matrix B_k is invertible (since all eigenvalues are negative) and we can rearrange Eq.(3) to

$$\boldsymbol{x}_k^* = -B_k^{-1}\boldsymbol{y}_k. \tag{5}$$

The matrix $-B_k$ is a non-singular M-matrix (non-positive off-diagonal entries and all eigenvalues with positive real part), hence its inverse is a positive matrix [5]. In particular, $\boldsymbol{x}_k^*$ is a positive vector unless $\boldsymbol{y}_k = \boldsymbol{0}$, in which case $\boldsymbol{x}_k^* = \boldsymbol{0}$. All in all, *a sub-critical SCC is either vanishing or positive.*

Now we can show **Condition 2**, by contradiction. Suppose that there are no critical SCC. If B_1 is super-critical, it must be vanishing. If B_1 is sub-critical, $\det(B_1) \neq 0$ and $B_1\boldsymbol{x}_1^* = \boldsymbol{0}$ (from Eq.(1)) implies $\boldsymbol{x}_1^* = \boldsymbol{0}$. Suppose now that all SCC B_l for $l < k$ are vanishing. Then either B_k is super-critical and hence vanishing, or B_k is sub-critical and hence also vanishing, by Eq.(5),

$$\boldsymbol{x}_k^* = -B_k^{-1}\left(\sum_{l<k} C_{kl}\boldsymbol{x}_l^*\right) . \tag{6}$$

Hence, if there are no critical SCC, then all $\boldsymbol{x}_k^* = \boldsymbol{0}$, and thus $\boldsymbol{x}^*$ is vanishing, which is in contradiction to the existence of a non-vanishing steady state $\boldsymbol{x}^* \neq 0$.

Case $\mu_k = 0$. In this case (critical SCC), Eq. (4) becomes, $\boldsymbol{v}_k\boldsymbol{y}_k = 0$. Since $\boldsymbol{y}_k \geqslant 0$ and $\boldsymbol{v}_k > 0$, Lemma 1 requires that,

$$\boldsymbol{y}_k = \sum_{l<k} C_{kl}\boldsymbol{x}_l^* = \boldsymbol{0}. \tag{7}$$

This reduces Eq.(3) to $B_k\boldsymbol{x}_k^* = \boldsymbol{0}$, so either $\boldsymbol{x}_k^*$ is zero or it is a non-negative 0-eigenvector of B_k, that is, a positive (Perron-Frobenius) eigenvector of B_k. We conclude that *every critical SCC is either vanishing, or positive.*

All in all, we have shown that *every SCC must be either vanishing, or positive.* If that is the case, $C_{kl}\boldsymbol{x}_l^* = \boldsymbol{0}$ implies $C_{kl} = \boldsymbol{0}$ or $\boldsymbol{x}_l^* = \boldsymbol{0}$ (if $\boldsymbol{x}_l^*$ is not zero, it must be positive, and the product of a non-negative matrix and a positive vector is zero if and only if the matrix is zero). In particular, $\boldsymbol{y}_k = \boldsymbol{0}$ if and only if $C_{kl} = \boldsymbol{0}$ or $\boldsymbol{x}_l^* = \boldsymbol{0}$ for all $l < k$. In words, $\boldsymbol{y}_k$ *is zero if and only if every SCC immediately upstream of* B_k *is vanishing.* (We call a SCC B_l immediately upstream of B_k if there is at least one link from a node in B_l to a node in B_k, that is, if $C_{kl} \neq \boldsymbol{0}$, the zero matrix.)

We now revisit the three cases above. If B_k is super-critical, we showed that $\boldsymbol{x}_k^* = \boldsymbol{0}$ and that $\boldsymbol{y}_k = \boldsymbol{0}$, that is, *all super-critical SCC must be vanishing, and all SCC immediately upstream of a super-critical SCC must also be vanishing.* If B_k is sub-critical, we showed that $\boldsymbol{x}_k^* = -B_k^{-1}\boldsymbol{y}_k$, with $-B_k^{-1}$ a positive matrix, so that $\boldsymbol{x}_k^* = \boldsymbol{0}$ if and only if $\boldsymbol{y}_k = \boldsymbol{0}$. That is, *a sub-critical SCC is vanishing if and only if every SCC immediately upstream is vanishing, otherwise it is positive.* Finally, if B_k is critical, we showed that $\boldsymbol{y}_k = \boldsymbol{0}$, that is, *all SCC immediately above a critical SCC must be vanishing.* Since, in addition, any SCC immediately upstream of a super-critical, or a vanishing sub-critical, SCC are vanishing, we have that *any SCC upstream of a critical SCC must be vanishing.* This, in particular, proves **Condition 3** and the last part of the theorem, and we are done. □

Next, we translate these mathematical theorems to the biological context of the Main Text.

- The equation $\frac{d}{dt}\boldsymbol{n}(t) = A\boldsymbol{n}(t)$ in the Main Text, Eq. **[5]** in Box 1 (whereby A may depend on $\boldsymbol{n}$), models the dynamics of the mean cell numbers $n_i(t)$ of cell state i at time t. The matrix A is a Metzler matrix for all $\boldsymbol{n}$, since the transitions rates $\kappa_{i\to j} \geqslant 0$ are non-negative. The graph $G(A)$ is the *cell state network*. Therefore, Theorem 3 applies. If the system parameters are independent of cell numbers $\boldsymbol{n}$, the system is linear and the stricter version, Theorem 1, applies.

- A directed path in the graph $G(A)$ corresponds to a *cell state trajectory*. The graph $G(A)$ therefore contains all possible cell state trajectories.

- According to the definition in the Main Text, an SCC in $G(A)$, characterised by reversibility of cell state trajectories, i.e. cyclic trajectories, corresponds to a *cell type*. Hence, we associate a cell type T_k with each SCC S_k $(k = 1, ...h)$ of $G(A)$.

- *Homeostasis* is a non-negative and non-vanishing steady state of cell numbers $\boldsymbol{n} = (n_1, n_2, ...)$, called $\boldsymbol{n}^*$, following dynamics according to Eq. **[5]** Main Text, i.e. $\frac{d}{dt}\boldsymbol{n} = 0$ at $\boldsymbol{n} = \boldsymbol{n}^*$.

- The dominant eigenvector μ_k of the adjacency matrix A_k of the SCC S_k is the *growth parameter* of cell type T_k. It determines the cell dynamics: if $\mu_k > 0$ then $|\boldsymbol{n}|$ increases asymptotically, if $\mu_k < 0$ then $|\boldsymbol{n}|$ decreases asymptotically, and if $\mu_k = 0$ then the mean cell number stays constant over time. We use this behaviour to define cell types: if $\mu_k > 0$, we call the cell type T_k *hyper-proliferative* (corresponding to a super-critical SCC), if $\mu_k < 0$ then T_k is *transient* (corresponding to a sub-critical SCC), and if $\mu_k = 0$ then T_k is a *self-renewing* cell type (corresponding to a critical SCC)[2].

Since matrix A in Eq. **[5]** in the Main Text (Box 1) is a Metzler matrix, and a homeostatic state $\boldsymbol{n}^*$ a non-negative non-vanishing steady state, Theorem 3 applies in this biological context, which translates to the following statements.

If a tissue cell population is in homeostasis then:

1. There is no hyper-proliferative cell type.

2. There is at least one self-renewing cell type.

3. Two self-renewing cell types are never on the same cell trajectory (i.e. connected by a directed path).

(If the system parameters are independent of the cell numbers $\boldsymbol{n}$, then the three conditions above are not only necessary, but also sufficient, by Theorem 1.)

Moreover, from the last part of Theorem 3 (or from Theorem 2 if the system parameters are independent of the cell numbers), it follows that

4. Any cells of a type upstream of a self-renewing cell type in the lineage hierarchy must vanish in homeostasis (they may, however, be present in non-homeostatic situations, such as development or regeneration).

We have, in particular, that the following must hold in any homeostatic renewing tissue.

- (From 4.) Self-renewing cell types must always be at the beginning of a homeostatic trajectory, i.e. at the apex of the homeostatic lineage hierarchy. Conversely, each apex of a lineage hierarchy must be a self-renewing cell type. In the Main Text, we define those cell types as *adult stem cells*. We note that multiple apices of a lineage hierarchy may exist.

[2] In case of an inert tissue, $\mu_k = 0$ could also denote *inert* cell types, however, throughout this work we only consider renewing tissues, and thus $\mu_k = 0$ must involve proliferative activity. Proliferative activity without net gain or loss of cells is self-renewing.

- When an adult stem cell divides, one or two of its daughter cells may be stem cells of the same type, but not of another stem cell type (otherwise, this would contradict Condition 3 above).
- Daughter cells of committed (progenitor) cells cannot be adult stem cells (this would contradict Condition 3 above).

We note that only for tissue cell systems where dynamic parameters are independent on the cell numbers, the stated conditions are also sufficient to ensure not only the existence but the stability of homeostasis [1]. If parameters depend on cell numbers, the system is non-linear, and we cannot draw conclusions about stability without knowledge of the particular dependence of parameters on cell numbers, that is, the underlying feedback regulation. In the following section, and in section 2C in the Main Text, we study the particular scenario of *crowding feedback*, and under which conditions such feedback assure a long-term homeostatic state.

2 Regulation of stemness by crowding feedback

We now consider the situation that cells are able to sense the cell density in their local environment – commonly called the 'niche' – and adjust their dynamic behaviour accordingly. In our model, this means that the kinetic parameters $\lambda_i, \omega_{ij}, d_i$ $(i, j = 1, 2, ..., m)$ – which in the following we denote simply as $\alpha_j, j = 1, 2, ..., 2m + m^2$ – depend on the number of cells of a certain type. Let us consider the dynamics of a cell type T, which resides at the apex of the lineage hierarchy, and we assume *crowding feedback*, i.e. that parameters depend only on the number of cells of type T, $N = \sum_{i \in T} n_i$ (see also the discussion in Main Text). Since the elements a_{ij} of the matrix A in Eq. 6, Main Text, depend on the parameters α_j, $a_{ij} = a_{ij}(\alpha_1, \alpha_2, ...)$, they also depend on N. Hence the dynamics are governed by the differential equation

$$\frac{d}{dt}\boldsymbol{n}(t) = A(N(t))\,\boldsymbol{n}(t) \tag{8}$$

where here, $\boldsymbol{n} = (n_1, ..., n_m)$ refers only to cell numbers of type T (m is the number of states of type T) and A is the sub-matrix for the dynamics of type T according to Eq. **[6]** in the Main Text[3]. The dependence $A = A(N)$ means that the elements of the matrix A are functions of N, and therefore also the dominant eigenvalue of A (the *growth parameter* in the Main Text) is also a function of N, $\mu = \mu(N(t))$, which thus becomes a dynamic quantity.

A self-renewing state is a particular cell configuration $\boldsymbol{n}^*$ with $N^* = \sum_{i \in T} n_i^*$ such that $\mu(N^*) = 0$. In the following, we consider also the slightly weaker definition of a *dynamic*, long-term self-renewing state, which prevails if the cell number does not increase or decrease in the long-term, i.e. the cell number may temporarily increase or decrease, but the cell numbers remain *on average* constant in the long term. This is the case if deviations from N^* are bounded. Such a state can be an asymptotically stable state at $\boldsymbol{n}^*$ or could involve fluctuations or oscillations around N^*, but no persistent deviation from N^* (see section 2C and Box 3 in the Main Text).

In the following, we show that a tissue cell population following the dynamics of Eq. (8) approaches a dynamic long-term self-renewing state (i.e. its asymptotic solutions are confined around the self-renewing state with $\mu = 0$) if,

$$\frac{\partial \mu}{dN} < 0 \text{ for all } N \geqslant 0\ , \tag{9}$$

with the additional assumption that all parameters, $\alpha_i, i = 1, 2, ...$, which may depend on N, converge for large N, i.e. $\alpha(N) \to \alpha_i^\infty$ for $N \to \infty$, to some limiting value α^∞. This is biologically reasonable as it simply states that cell processes cannot become infinitely fast.

[3]Note that here, for convenience, we changed the notation compared to the Main Text and to Section 1 of this document.

In order to show that condition (9) is sufficient for long-term self-renewal, we study what happens if under these conditions the solutions to Eq. (8) were not confined. In that case, the cell number N would either diverge or decay to zero.

First, let us assume that N diverges, i.e. $N \to \infty$ for $t \to \infty$. In that case, all parameters α_j attain their limiting values α_∞ and the matrix A becomes essentially constant, $A \to A_\infty$, whereby $A_\infty = [a_{ij}(\alpha_1^\infty, \alpha_2^\infty, ...)]_{i,j}$. In that case, the dynamical system (8) becomes linear, and a linear system converges to its dominant eigenvector for $t \to \infty$: $\boldsymbol{n} \to e^{\mu_\infty t}\hat{\boldsymbol{n}}_\infty$, where μ_∞ is the dominant eigenvalue of A_∞ and $\hat{\boldsymbol{n}}_\infty$ its normalised, positive dominant eigenvector. Notably, since T is a cell type, its cell state network is strongly connected, hence A_∞ is irreducible with non-negative off-diagonal elements, i.e. a Metzler matrix. The Perron-Frobenius theorem hence assures that the dominant eigenvalue μ is simple. Now, since according to condition (9), $\frac{\partial \mu}{dN} < 0$ for all N, we have that $\mu < 0$ for $N > N^*$, hence $\mu_\infty < 0$ and the system will for large N converge to $e^{-|\mu_\infty| t}\hat{\boldsymbol{n}}_\infty$. Thus, N will asymptotically decrease. This is in contradiction with our assumption that $N \to \infty$.

Now let us assume that $N \to 0$ for $t \to \infty$. That means that $\boldsymbol{n} = \boldsymbol{0}$ is a stable fixed point of the dynamical equation (8). Then, the dominant eigenvalue of the Jacobian matrix at $\boldsymbol{n} = 0$ has a negative dominant eigenvalue. The Jacobian matrix is,

$$J = [J_{ij}] \text{ with } \tilde{J}_{ij} = \frac{\partial(\frac{d}{dt}n_i)}{\partial n_j} = \frac{\partial(A\boldsymbol{n})_i}{\partial n_j} = a_{ij} + \sum_k \frac{\partial a_{ik}}{\partial n_j} n_k \tag{10}$$

Notably, at $\boldsymbol{n} = 0$, we have $n_k = 0$ for all k and thus $J = A$. In particular, the dominant eigenvalue of J is that of A, i.e. $\mu(N = 0)$. However, since $\frac{\partial \mu}{\partial N} < 0$ everywhere, we have that $\mu > 0$ at $N = 0$ and thus the fixed point at $\boldsymbol{n} = 0$ is unstable, contradicting our assumption.

Thus we have shown that the cell number N can neither diverge nor approach zero if (9) is fulfilled. Furthermore, since $\frac{\partial \mu}{\partial N} < 0$ everywhere, the function $\mu(N)$ is monotonic and thus there cannot be any other non-zero fixed point with $\mu(N) = 0$. Therefore, the solutions to Eq. (8) are confined around $\boldsymbol{n}^*$, if condition (9) is fulfilled.

Hence, condition (9) is a sufficient condition that $\boldsymbol{n}$ either converges to $\boldsymbol{n}^*$, with $\mu(\boldsymbol{n}^*) = 0$, or to a trajectory that is confined around it. In particular, the cell numbers of cell type T do not, in the long term, increase or decrease. They may, however, oscillate or fluctuate around this value, i.e. temporarily one may have $\mu > 0$ or $\mu < 0$, corresponding to temporary configurations with $N > N^*$ and $N < N^*$, but cell numbers always return towards N^*, so that N does not deviate substantially from N^* in the long run, and the long-term average stays constant over time. We can therefore refer to the resulting state as a dynamic *long-term* self-renewing state. As the cell type T is at the apex of a lineage hierarchy, this means that T is a stem cell type in the long term. From condition (9) also follows a sufficient sign condition for homeostasis: if all terms $\frac{\partial \mu}{\partial \alpha_l}\frac{\partial \alpha_l}{\partial N}$ are negative, which is fulfilled if the sign of $\frac{\partial \alpha_l}{\partial N}$ is opposite the sign of $\frac{\partial \mu}{\partial \alpha_l}$ for all *relevant* parameters α_l (those parameters with $\frac{\partial \mu}{\partial \alpha_l} \neq 0$), then

$$\frac{d\mu}{dN} = \sum_i \frac{\partial \mu}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial N} < 0 \ . \tag{11}$$

Hence, a dynamic self-renewing state, with constant cell numbers in the long-term average, is assured if all relevant parameters have feedback with the sign of $\frac{\partial \alpha_i}{\partial N} = \alpha_i'(N)$ opposite to the sign of $\frac{\partial \mu}{\partial \alpha_i}$. This is precisely the sign criterion of Box 3 in the Main Text.

To demonstrate how a cell type becomes self-renewing by crowding feedback, we studied two numerical example scenarios in Figure 1. The figure shows how a particular cell type T_a, implemented as a cell state network forming a single SCC (depicted in the top panels), attains a (dynamic) self-renewing state if the sign condition for crowding feedback (Eq. (9) and Box 3, Main Text) is fulfilled. It shows that for various initial conditions, which might be not self-renewing, the growth parameter approaches or oscillates around the value $\mu_a = 0$ (bottom panels). Consistently, the cell number N_a asymptotically

Case	A							B			
α	λ_1	λ_3	ω_{13}	ω_{23}	ω_{21}	δ_1	δ_3	λ_1	λ_2	ω_{12}	ω_{21}
$\alpha^{*\prime}$	< 0	< 0	< 0	< 0	> 0	> 0	> 0	< 0	< 0	< 0	> 0
c	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.100	0.005	0.001	1.000
k	1.02	1.01	1.12	1.85	1.52	2.68	1.64	0.68	13.54	0.18	0.87
n	2	2	2	2	2	2	2	50	2	5	1

Table 1: Values of the Hill function parameters used to describe the kinetic parameters in the numerical example of crowding-dependent regulation of stemness. The corresponding networks are shown in the top panels of Figure 1. The generic kinetic parameter which is function of the total number of cell N_a is given by $\alpha(N_a) = c + k(N_a/N)^n/(K^n + (N_a/N)^n)$ when $\alpha^{*\prime} > 0$ and $\alpha(N_a) = c + k/(K^n + (N_a/N)^n)$ when $\alpha^{*\prime} < 0$. Common values are $K = 1$ and $N = 10^3$. Arbitrary time units are assumed in both examples. In the asymptotically stable example, indicated as case A, the following parameter values to define the system are also considered: $r_{11} = 1$, $r_{12} = 1$, $r_{13} = 0$, $r_{31} = 0$, $r_{32} = 2$ and $r_{33} = 0$. Additional parameters for the locally unstable example, case B are: $\delta_1 = 0.92$, $r_{11} = 0.09$, $r_{12} = 0.11$, $r_{21} = 0.47$ and $r_{22} = 0.53$ (parameter names as in Box 1 in the Main Text)

remains confined around the self-renewing state (middle panels), corresponding to a dynamic homeostatic state. The two examples are representative of an asymptotically stable system (case A, left panels) and a system with stable oscillations (case B, right panels). Whilst in the asymptotic stable case the steady state is reached and maintained, in the oscillating case the number of cells keeps oscillating periodically around the self-renewing condition, but it does neither diverge nor decay to zero.

In order to fulfill condition (9) for crowding feedback in the examples, we chose each kinetic parameter $\alpha_i \in \{\lambda_j, d_j, \omega_{jk}\}_{j,k=1,\dots,m}$ being a function of N_a. In particular, we chose a Hill function [6] of the type $\alpha_i(N_a) = c_i + k_i(N_a/N_i)^{n_i}/(K_i^{n_i} + (N_a/N_i)^{n_i})$ in case α_i is an increasing function of N_a and $\alpha_i(N_a) = c_i + k_i/(K_i^{n_i} + (N_a/N_i)^{n_i})$ in case of a decreasing function of N_a. To find suitable models (i.e. such that $\mu_a(N_a^*) = 0$ and the real part of the eigenvalues of the corresponding Jacobian matrix are all negative for the asymptotically stable case and at least one positive in the locally unstable case with oscillations), we implemented a random search employing a genetic search algorithm [7]. More specifically, we first searched for a set of values α_i^* for which μ_a, is zero. Then, we tested the stability of random values $\alpha_i^{*\prime} = \frac{\partial \alpha_i}{\partial N_a}$, for which the sign criterion is satisfied (i.e. $\text{sign}(\alpha_i^{*\prime}) = -\text{sign}(\frac{\partial \mu_a}{\partial \alpha_i})$). Once the model is derived, the parameters of a Hill function consistent with α_i^* and $\alpha_i^{*\prime}$ are determined, considering also some upper and lower saturation levels for the kinetic parameters. Two cases, one asymptotically stable and one oscillating (which is locally unstable at the fixed point $\boldsymbol{n}^*$), are picked to illustrate the dynamics. The values of the model parameters, assuming arbitrary units, are reported in Table 1.

We then solved the differential equations describing this system, Eq. (8), numerically for three sets of initial conditions corresponding to 0.8, 1 and 1.2 times the expected steady state $\boldsymbol{n}_a^*$. The solutions, indicated respectively as $\boldsymbol{n}_a(t = 0) = 0.8\boldsymbol{n}_a^*$, $\boldsymbol{n}_a(t = 0) = \boldsymbol{n}_a^*$ and $\boldsymbol{n}_a(t = 0) = 1.2\boldsymbol{n}_a^*$ are plotted in terms of $N_a(t)$ (middle panels) and $\mu_a(t)$ (bottom panels). We observe that for the not self-renewing initial conditions in the asymptotically stable case A, the system eventually attains a steady state $N_a \to N_a^*$ (middle-left panel) and self-renewal property $\mu_a \to 0$ (bottom-left panel) over time t. For the locally unstable case B, a phase plot is also shown (insert in the middle-right panel) confirming that the number of cells converge to a cyclic trajectory around the self-renewing condition, and thus remain confined.

References

[1] P. Greulich, B. D. MacArthur, C. Parigini, and R. J. Sánchez García. Stability and steady state of complex cooperative systems: a diakoptic approach. *Royal Society Open Science*, 6:191090, 2019.

[2] M. W. Hirsch and H. Smith. Monotone dynamical systems. *Handbook of differential equations: Ordinary Differential Equations*, page 57, 2006.

[3] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

[4] K. J. Arrow. A "dynamic" proof of the Frobenuis-Perron theorem for Metzler matrices. *Probability, Statistics, and Mathematics*, pages 17–26, 1989.

[5] C. D. Meyer and M. W. Stadelmeier. Singular M-Matrices and Inverse Positivity. *Linear Algebra and Its Applications*, 22, 1978.

[6] J. Lei, S. A. Levin, and Q. Nie. Mathematical model of adult stem cell regeneration with cross-talk between genetic and epigenetic regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(10), 2014.

[7] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.

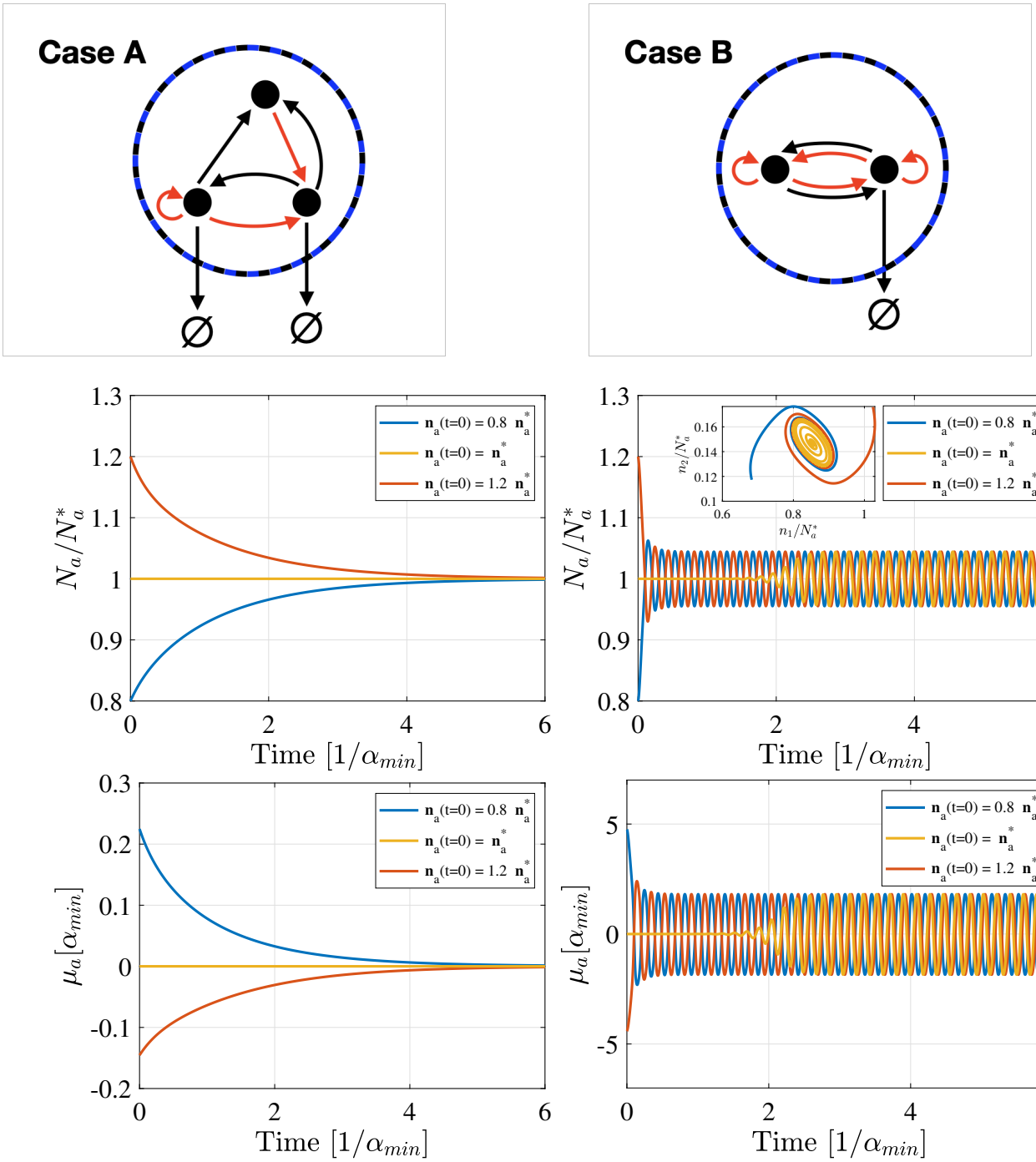

Figure 1: Example of crowding-dependent regulation of self-renewal capacity for a cell type T_a with 3 and 2 cell states (shown as cell state networks on the top panels and indicated respectively as case A and B. Symbols and colours are as in Fig. 3 of the Main Text). For each kinetic parameter α_j we chose the crowding dependence based on a Hill function such that the criterion, Eq. (9) (and Box 3 in the Main Text), is fulfilled (see text for how the function parameters are found and Table 1 for the specific values). The middle panels show the time evolution of mean cell numbers N_a (obtained as numerical solutions of the dynamic equation Eq. (8) for type T_a), for different sets of initial conditions: $\boldsymbol{n}_a(t=0) = \boldsymbol{n}_a^*$ corresponds to an initially self-renewing case; $\boldsymbol{n}_a(t=0) = 0.8\boldsymbol{n}_a^*$ and $\boldsymbol{n}_a(t=0) = 1.2\boldsymbol{n}_a^*$ are instead representative of initially not self-renewing conditions. The bottom panels similarly show the time evolution of the growth parameter μ_a. Time is scaled by the inverse of the smallest rate $\alpha_{min} = \min_{i,j}\{\lambda_i, \omega_{i\to j}, d_i\}$ at N_a^*. Crucially, we observe (in bottom panels) that even if T_a is not initially self-renewing ($\mu_a(t=0) \neq 0$) the parameters self-adjust over time such that the growth parameter attains $\mu_a = 0$ (case A, bottom-left) or oscillates around this condition (case B, bottom-right). This is also reflected in the time evolution of the cell number, which becomes stationary, $N_a \to N_a^*$ (case A, middle-left panel), or oscillatory with constant amplitude around N_a^* (case B, middle-right panel).