Trends in Ecology & Evolution

CellPress

Review

Evolution of Cellular Differentiation: From Hypotheses to Models

Pedro Márquez-Zacarías,[1,2,@] Rozenn M. Pineau,[1,2,@] Marcella Gomez,[3] Alan Veliz-Cuba,[4] David Murrugarra,[5,@] William C. Ratcliff,[2,@] and Karl J. Niklas[6,*]

Cellular differentiation is one of the hallmarks of complex multicellularity, allowing individual organisms to capitalize on among-cell functional diversity. The evolution of multicellularity is a major evolutionary transition that allowed for the increase of organismal complexity in multiple lineages, a process that relies on the functional integration of cell-types within an individual. Multiple hypotheses have been proposed to explain the origins of cellular differentiation, but we lack a general understanding of what makes one cell-type distinct from others, and how such differentiation arises. Here, we describe how the use of Boolean networks (BNs) can aid in placing empirical findings into a coherent conceptual framework, and we emphasize some of the standing problems when interpreting data and model behaviors.

Highlights

The evolution of multicellular life required the emergence and evolution of multiple cell types.

Cell types can evolve via different mechanisms, including ecological context, genomic innovation, cooperative integration, or a combination of these.

Studying cell types in multicellular organisms has been recently enhanced by new technologies that allow for single-cell profiling.

Despite these advances, the interpretation of data and the concept of 'cell types' remain ambiguous.

New theoretical work is needed to put the empirical and conceptual advances into a coherent framework.

Boolean networks are models that can bridge the conceptual and theoretical issues with the empirical findings.

Multicellular Evolution and Biological Complexity

The evolution of multicellularity was the **evolutionary transition** (see Glossary) responsible for the most dramatic and ecologically important increases in **organismal size** and complexity [1,2]. A prominent feature observed in most **multicellular organisms** is that they contain a diversity of **cell-types**, which allows for the performance of complex behaviors and functions [3–5]. Recent technological advances in **single-cell profiling** have allowed the mapping of cell diversity at an exquisite resolution within individual multicellular organisms [6–8], but little theoretical work has followed these empirical observations and, as a result, the concept of 'cell-type' remains ambiguous. Thus, a key goal of current research on multicellular evolution is to bridge these rapid developments in empirical research with theoretical and conceptual advances. Here, we describe the problem of explaining cellular differentiation in multicellular organisms, and show how **Boolean networks (BNs)**, a class of discrete dynamical systems, provide a path forward for solving several major outstanding questions.

Multicellular Diversity

Multicellularity has evolved independently in many different **lineages**, and has been experimentally induced and evolved *in vitro*, indicating that it can be stabilized under a varied set of ecological and evolutionary conditions [4,9–12]. Although there are diverse lineages of multicellular organisms across the tree of life, taxa can be assigned to one of two groups: clonal, which are intrinsically multicellular (e.g., land plants and metazoans), or aggregative, which facultatively form multicellular groups (e.g., slime molds) [13]. In **clonal multicellularity** (Figure 1A), organisms typically develop from a single cell (a **zygote**, an asexual spore, or a propagule of clonal cells) and result in organisms with genomically highly related cells that can achieve functionally distinct cell-types. **Aggregative multicellularity** (Figure 1B) develops when independent free-living unicellular organisms 'come together' and form a new organism that contains potentially genomically unrelated cells [14,15]. Intrinsically multicellular organisms develop from single cells (or single-cell bottlenecks), as opposed to aggregative multicellularity. Clonal multicellular organisms achieve **intercellular cooperation** more readily and manifest more cell types compared with aggregative multicellularity

[1]Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, GA, USA
[2]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[3]Department of Applied Mathematics, University of California Santa Cruz, Santa Cruz, CA, USA
[4]Department of Mathematics, University of Dayton, Dayton, OH, USA
[5]Department of Mathematics, University of Kentucky, Lexington, KY, USA
[6]School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

*Correspondence: kjn2@cornell.edu (K.J. Niklas).
@Twitter: @PedroM_Z (P. Márquez-Zacarías), @PineauRozenn (R. Pineau), @DiscreteMathBio (D. Murrugarra), and @wc_ratcliff (W.C. Ratcliff).

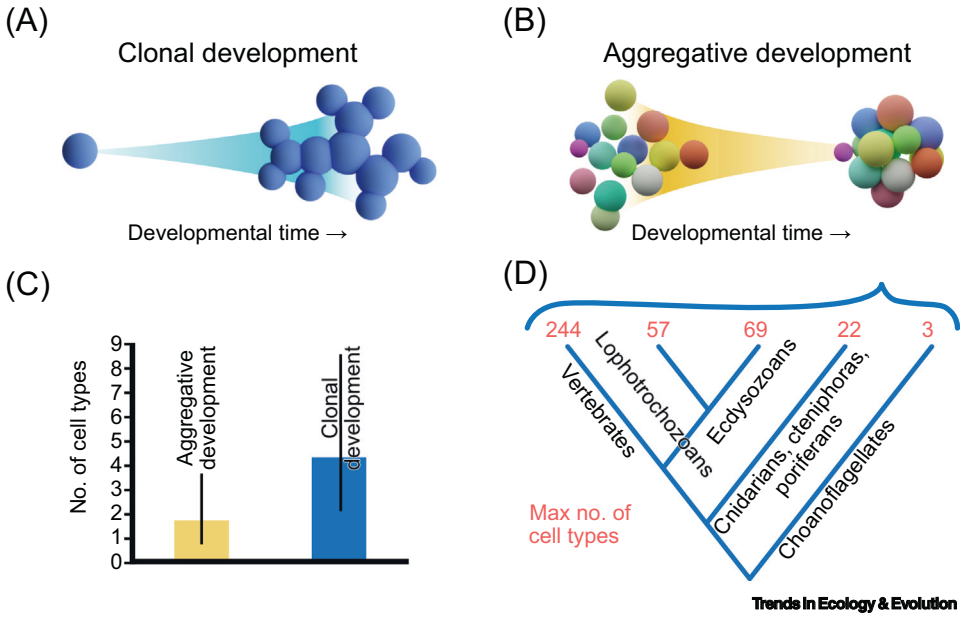

Figure 1. Key Differences between Aggregative and Clonal Multicellularity. Clonal multicellular organisms develop from a single-cell bottleneck (A) and, thus, the cells within the fully developed organism have the same genetic identity (they are 'clones'). By contrast, aggregative multicellular organisms develop via aggregation of individual cells (B), leading to a chimeric group. It has been described that clonal multicellularity exhibits a higher number of cell-types compared with aggregative multicellularity (C). Another important trend within clonal multicellularity is that the more complex the organisms are, the more cell-types they have (D). Adapted from [5] (C) and [71] (D).

(Figure 1C,D) [5,16,17]. Additionally, multicellular organisms of larger size tend to have more cell-types, a trend known as the 'size–complexity rule' [3] (Box 1). However, some of these trends likely reflect a historic over-representation of animals in data sets. Thus, more data are needed from groups other than animals (for exceptions, see [16,18,19]).

The stark difference in the number of cell types between aggregative and clonal organisms lacks a clear explanation. Clonal **development** explains some features of complex multicellularity, such as the reduction of **within-organism conflict** [9] and the increase in organismal size, but the lack of conflict itself cannot drive the evolution of cell types, since even aggregative, potentially chimeric, organisms can develop into distinct cell-types [17,20]. Additionally, the size–complexity rule cannot account for the fact that even unicellular organisms manifest different cell types during their life cycles (however, for a recent theoretical approach, see [21]). The fact that multiple cell-types exist within unicellular and aggregative organisms suggests that the origin of cellular differentiation is largely independent of the route to multicellularity [14,15]. Thus, the task at hand becomes finding an explanation for cell differentiation in any given multicellular context, being clonal or aggregative, without conflating explanations for multicellular evolution itself.

What Counts as a Cell Type?

A meaningful explanation of the evolution of cell types requires *de minimus* a clear and canonical definition of what counts as a 'cell-type'. At first glance, it would appear that cell types differ in terms of either their morphology or function, or both. Surprisingly, however, there is no general agreement on the extent to which morphological or functional divergence is required to identify a particular type of cell as different from other types, and the answers largely depend on the particular organism of interest, or a given method of classification (for a recent survey on cell-type definitions, see [22]).

Glossary

Aggregative multicellularity: a group of cells comprising nonclonal cells, developed via the accumulation of formerly independent (free living) cells.
Boolean network (BN)/Boolean model: a discrete dynamical system comprising a set of Boolean variables (e.g., True/False, 0/1), the values of which depend on a set of functions applied to each variable.
Cell type: a cell population phenotypically, functionally, or morphologically different from any other in the same organism.
Clonal multicellularity: an organism comprising many cells derived from a single cell or propagule and, thus. composed of genetically identical cells.
Cyclic attractor: the solution of a BN can take the form of an orbit, a closed trajectory in the phase space that always returns to the starting point.
Development: formation of a multicellular organism via either clonal or aggregative mechanisms.
Division of labor: capacity for components of a system to perform different functionalities aligned with a system-level aim.
Dynamical patterning modules: set of genetic toolkits and physical motifs in a multicellular system allowing for its development.
Evolutionary transitions: framework to understand the evolution of complex life. When it highlights the changes from a group of previously independent individuals, to the composition of new and more complex entities, this refers to evolutionary transitions in individuality.
Functional heterogeneity: one cell/different cells can have a palette of different behaviors according to several variables: the phase of the cell cycle, the environment, interactions with other cells, and so on.
Intercellular cooperation: combined effort of different types of cell to achieve the same goal (e.g., colonization of a new environment).
Lineage: line of descendants from a particular ancestor.
Morphogenesis: process during which the structure of an organism shapes.
Multicellular organisms: entities comprising two or more cells, as opposed to unicellular organisms.
Ontogeny: entire sequence of the events shaping the developmental

Box 1. Size–Complexity Rule

The size–complexity rule simply summarizes the observation that the largest organisms tend to be more complex (Figure I). Complexity in this case can be defined as the number of cell-types (Figure I). This trend is sometimes associated with the evolution of the division of labor, because different cell-types represent distinctive tasks that are functionally integrated into a single individual. The key observation is that the biomass (in number of cells) defines the number of cell-types an organism exhibits. However, changing the number of cell-types does not affect the size of an organism, which suggests that organismal size is causally linked to the complexity level of an organism [1,3,71]. Over the past few decades, the increased collection of size and complexity data over a range of taxa has allowed a finer empirical verification of this relationship. For instance, if we classify all living entities on 'complexity groups' (see [1] for details), we observe that the organismal size distribution is unimodal. This means that the size increase between groups is more important than within-groups, which furthermore suggests a lower and upper boundary of size in each complexity group.

Although widely observed across taxa, this rule is approximate and subject to bias toward some groups of organisms (e.g., animals tend to be over-represented). A theoretical understanding nuances the trend by showing that organismal shape is another important variable affecting size and complexity. The rule is violated by filamentous organisms (filamentous cyanobacteria), as opposed to more compact structures, such as spheroid-like organisms (volvocine green algae), and clonal multicellular organisms [21], indicating that the size–complexity trend depends on a more complex interplay than previously thought.

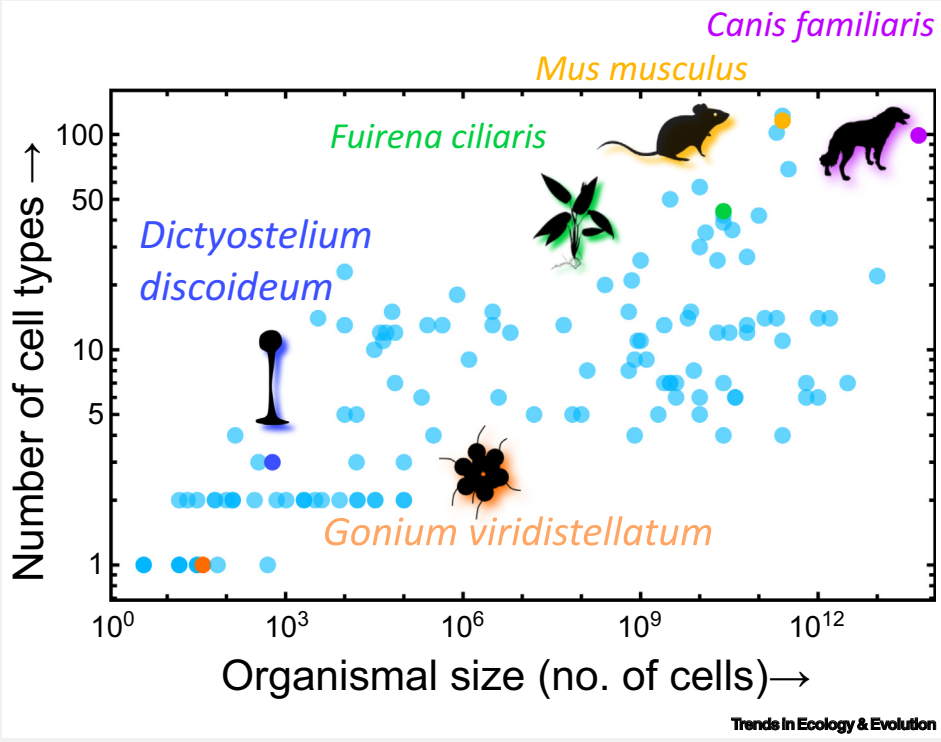

Figure I. The Size–Complexity Rule. It has been proposed that the size of an organism scales with its complexity. In this context, an approximation for complexity is the number of cell-types within a given organism. This conceptualization is supported by a large body of evidence, as shown in this plot, which highlights some representative multicellular groups, from simple to complex for example, the slime mold (*Dictyostelium discoideum*), green algae (*Gonium viridistellatum*), land plants (*Fuirena ciliaris*), mouse (*Mus musculus*), and dogs (*Canis familiaris*). Adapted from data in [5].

history of an individual organism over time.

Organismal size: quantitative measure of the size of an entity; can range, for example, from the number of cells in an organism, to its height or volume.

Point attractor: the solution of a BN can be a 'frozen' point, where the network remains in this state regardless of more model updates.

Separatrix: boundary that separates distinct modes of behavior in a dynamical system, for instance, attractors of a system.

Single-cell profiling: portrait of a single cell drawn with genetics, transcriptomic, or metabolomic tools (or any other phenotypic characterization, usually molecular).

Symmetry breaking: bifurcation-like phenomenon, usually the product of the amplification of differences from an initially homogeneous state.

Waddington's landscape: metaphor by Conrad H. Waddington where the cell is a ball, rolling on an uneven hill marked with slopes and valleys. The landscape represents the cell differentiation process, and its features will guide the ball to a stable position at the bottom of the hill.

Within-organism conflict: when cells arise from diverse genomic lineages, competition can occur between them, as in the case of aggregative multicellular organisms.

Zygote: first cell in multicellular and sexually reproducing eukaryotes, formed by the fusion of the gametes.

This ambiguity likely emerges because the definition of a cell-type simultaneously presents empirical, conceptual, and theoretical challenges that raise open-ended questions on each front. For example, there has been a progressive refinement of cellular and molecular techniques (e.g., single-cell RNA sequencing) aimed at empirically quantifying the molecular composition of single cells to distinguish cell types within multicellular organisms [7,23]. Indeed, the degree of technical

refinement appears to be progressing exponentially [8]. However, advances in molecular evaluations of cell-types have not been matched with equal advances in the theoretical or conceptual tools required to identify the thresholds that must be crossed for a particular cell to be classified as morphologically, physiologically, or functionally different from any other [22,24]. Nevertheless, amid the lack of conceptual clarity about what constitutes a cell-type is a recent framework by Arendt *et al.* that aims to distinguish cell-types within and between species [25].

Multiple theoretical and conceptual problems emerge when trying to interpret the wealth of available data, or when building mathematical models of cellular differentiation [26]. When interpreting data, complications arise because cells inherently behave in heterogeneous ways (even isogenic cells under the same environmental conditions), which, in some cases, may manifest as **functional heterogeneity** [27,28]. In addition, a single cell-type can appear to behave differently depending on how the data are analyzed. Moreover, this heterogeneity often comes from 'hidden variables' that cannot be directly accessed from the data themselves, making it hard to ascribe explanations to such variation [29]. Another source of disconnect between data and theory arises in the opposite extreme, a continuum of behaviors wherein variation sits on a 'continuous manifold' that is hard to partition precisely because it does not look suitable for discretization [30].

To exemplify how data interpretation might be problematic, consider recent studies at single-cell resolution (e.g., [6,23,31]). A basic assumption is that the data can be organized into 'clusters' that faithfully represent different cell types (Figure 2). However, most classification methods based on single-cell data require researchers to train models with labeled cells, or with some expected number of cell-types *a priori* (e.g., [32,33]). These nonobjective aspects of the analysis make the interpretation of the data ambiguous and hard to compare across experiments and analysis pipelines. All of these limitations highlight the fact that fundamental technical-empirical questions remain. For example, is each data cluster a true 'cell–type', or could they be snapshots of the dynamics of a single type? The opposite question also emerges: could different cell types be misassigned to the same cluster simply because they happen to have partially overlapping dynamics?

These limitations also open the opportunity to use theoretical approaches that do not rely heavily on data acquisition interpretation, although these present equally important challenges. Any theoretical approach must ultimately rest on empirically reliable observations (or, at the very least, be tested against solid data). In the case of cellular differentiation, a few models have been formulated over the decades. These include models based on continuous dynamical systems and related formalisms (e.g., [34]). However, by far the most influential have been BNs and logical models introduced by Kauffman and Thomas, respectively [35–37]. The main distinction between them is the discrete nature of BNs in both phase space (e.g., the set of all possible states that a dynamical system can take) and time, as opposed to the continuous character of continuous dynamical systems (e.g., differential equations). An advantage of BNs is that they do not require as many details as continuous models and, thus, can model qualitative behaviors readily, such as cellular differentiation.

The use of BNs (also referred to as Boolean models) has been primarily limited to the study of their dynamical properties and the modeling of the cellular differentiation process; thus, a lot of work remains to be done about the evolutionary aspects of this process. However, the potential of this formalism to tackle these aspects can only be appreciated after a revision of current hypotheses on the evolution of cellular differentiation.

Explaining how cell-types emerge and evolve in multicellular organisms requires one to consider many factors that interact in complex ways. Multicellular organisms achieve cellular diversity

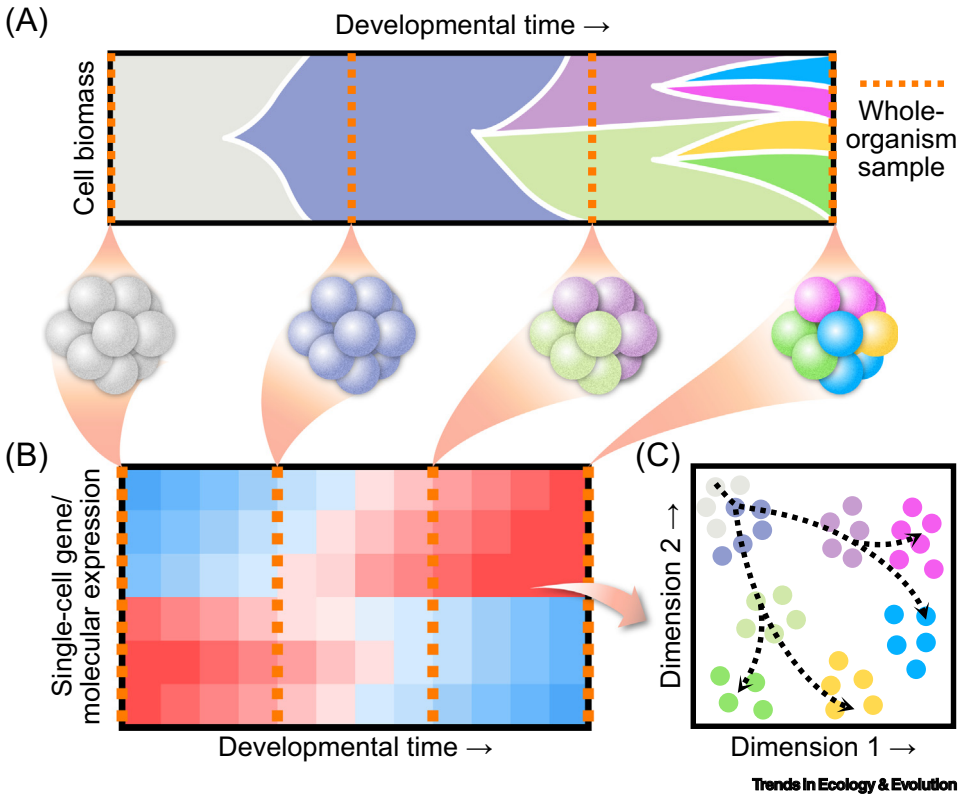

Figure 2. Methodological Process for Detecting Cell Differentiation in Multicellular Organisms. In recent years, there has been an explosion of technologies that allow for the study of multicellular organisms at single-cell resolution [23,24,29]. Although many details depend on the particular system of study and methodology of use, there are some common patterns in the methodologies. A developing multicellular organism can be studied at multiple phases of its development (A), ideally at the phases in which there is change in cell phenotypes. At these particular stages, the single cells of the organism are analyzed for phenotypic changes (B), usually by molecular profiling (e.g., gene expression, proteomic, or metabolomic analysis). Finally, from the single-cell profiles, it is common to apply a dimensionality-reduction process to the data. Finally, a clustering analysis will reveal those profiles that are most similar to each other (C), ideally corresponding to cell types.

throughout their **ontogeny**, but this diversity needs to be functionally integrated, so that no single given cell takes advantage of the system. Indeed, it has been proposed that the failure to maintain such functional integration in multicellular organisms underlies neoplasms and cancer [38,39]. Intercellular cooperation stabilizes an evolving system that is nevertheless flexible enough to allow for functional novelties, while preserving previous adaptations [26]. There are many hypotheses to explain how these processes can ultimately coexist in single multicellular organisms, and we highlight a few of them in Box 2, noting that our categorization is merely heuristic and by no means exhaustive.

Boolean Networks as Models for Cellular Differentiation

The evolution of cellular differentiation arose independently several times; thus, it is reasonable to think that abstract models can be more useful than detailed ones, because the mechanisms maintaining multicellularity differ widely among lineages (e.g., cell-to-cell adhesion is achieved in very different ways in plants and metazoans). BNs are a class of discrete dynamical systems initially developed to model cellular differentiation as a dynamical process (e.g., [35,36]). However,

the idea is not new, and was in fact inspired by the work of Conrad H. Waddington, who portrayed with his famous metaphorical landscape the developmental process, which included cellular differentiation [40] (see Figure ID, Key Figure in Box 2).

The Theoretical Basis of Boolean Networks

BNs are abstract constructs used to model cell differentiation without referring to a particular biological system [41]. One advantage of this discrete approach is that it allows one to ask concrete theoretical questions without focusing on the diverse details of a real biological system, which is required in most continuous models (e.g., with the use of differential equations), where empirically verified rate constants are usually necessary. Instead of considering the details of a system, BNs focus on the logic of the interactions between the components of the systems. Although they might appear abstract at first glance, BN models have been used to accurately model cellular differentiation in numerous systems, including plants, flies, yeast, humans, and many other kinds of organism (e.g., [42–47]). Thus, the Boolean framework allows one to pose specific questions and to test specific hypotheses in an intuitive way without losing power or generality (e.g., [48,49]). So far, no other model of cellular differentiation has the same advantages.

Box 2. Evolutionary Origins of Cellular Differentiation

Here, we briefly discuss some ideas on the evolution of cellular differentiation, noting that these are not necessarily exclusive or exhaustive.

Novel or Co-Opted Genes

It has been observed that novel or co-opted genes allow for the evolution of cell types [72–77], driving innovation in plants, flies, and other organisms [78–80] and the evolution of multicellularity itself [19,77,81,82] (Figure IA). Co-option is widespread, because the unicellular ancestors of multicellular lineages have machinery that readily allows for multicellularity to evolve [19,83–85]. Furthermore, organisms can expand their functional repertoire while preserving previously evolved cell types (e.g., [25,86]). Interestingly, adaptive changes fix at different rates between unicellular and multicellular organisms due to the distinct ways in which genetic variation is partitioned.

Division of Labor

Division of labor (i.e., task allocation) within a multicellular organism can drive cell differentiation (Figure IB). This has been observed in *Volvox* and cyanobacteria, among other organisms (e.g., [4,87–89]), and recent theoretical work helps explain when this process is favorable [90,91]. Additionally, microbial division of labor shows a similar functional pattern to multicellular organisms, leading to the interpretation that this mechanism is widespread and ancient [92–94].

Life-Cycle Integration

A recent hypothesis is that cell differentiation can emerge through the integration of the functionally distinct life stages of a unicellular organism (Figure IC). This hypothesis has been called 'temporal to spatial' evolution of cell differentiation [4,95], and stems from the fact that single-celled organisms can undergo various forms of differentiation though their life stages, as in *Dictyostelium* and some bacteria [16,62,96,97]. This hypothesis can be related to the idea of 'ecological scaffolding' [98] and symbiotic interactions [99].

Dynamical Systems Behavior

Another hypothesis is that cell types emerge naturally (and expectedly) from the dynamical properties of cells (Figure ID). This hypothesis has its roots in the concept of **symmetry breaking** in biological systems (e.g., Turing's model of **morphogenesis** [100]). The idea is that a system that is homogeneous at some point, can naturally become heterogenous over time (the symmetry is 'broken'). **Waddington's landscape** is a portrait of the temporal dynamics of a system undergoing symmetry breaking (Figure ID). In this landscape, the development of a zygote (a single cell type) results into a mature multicellular organism (with many cell-types).

Two additional hypotheses worth mentioning are the evolution of cell-types as a response to the adaptation to stress [64], and via phenotypic plasticity [65].

Key Figure

Four Hypotheses to Explain the Origins of Cell Differentiation

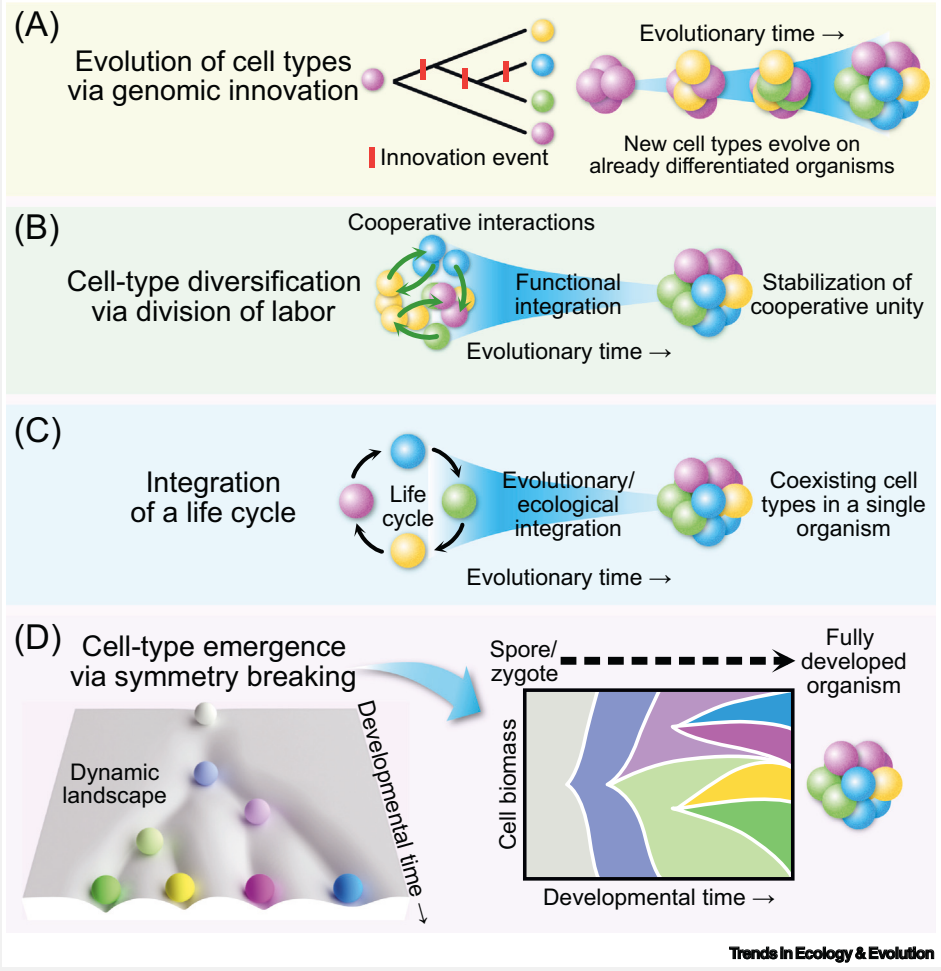

Figure I. Genomic innovation can drive the origins of novel cell-types while preserving previous innovations (A). A different hypothesis is that cooperative interactions via division of labor can stabilize the coexistence of differentiated cells into a single organism. Another idea proposes that the ecological context of differentiated cells, for instance in single-celled organisms with complex life cycles, can lead to the evolutionary integration of the differentiated cells (C). Finally, it has been proposed that the differentiation of cells in a multicellular organism follows from the 'symmetry breaking' occurring when an initially homogeneous system (e.g., the zygote) develops into multiple stable behaviors (D). This is a modern interpretation of Waddington's landscape [40].

BNs model cell differentiation and gene regulation by representing the states of genes (or other molecules) as either ON or OFF (see Figure 3 for visual overview of the main features of BN). However, this is not a limitation; more generally, multistate models allow genes to take on more than two states to incorporate more complex interactions [50]. To incorporate the inherent stochasticity of regulatory processes, different generalizations of BN have been developed, with the stochasticity originating from updating schedules [51]. Standard updating schedules include synchronous schedules, where all the nodes (representing genes or other participating

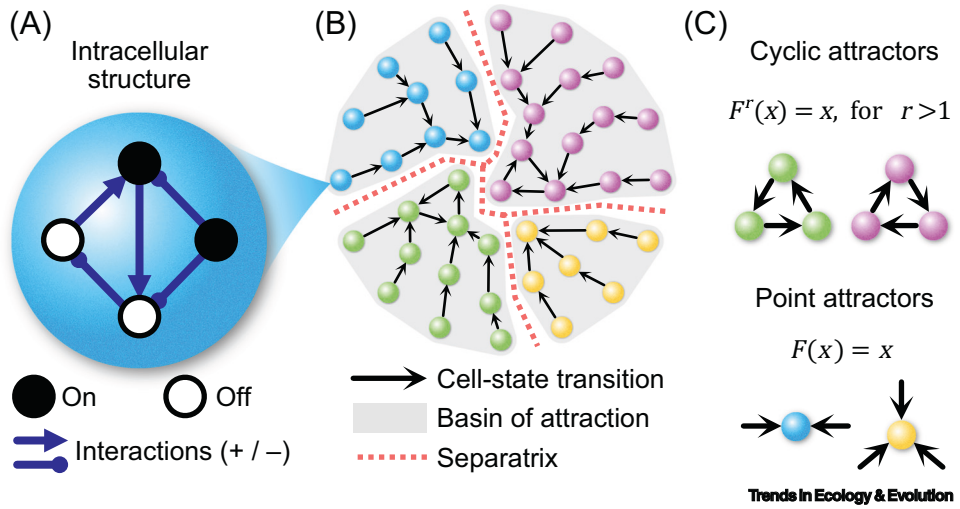

Figure 3. Modeling of Cell Differentiation with Boolean Networks (BNs). A BN can model the phenotypic dynamics of cells by representing their genes (or any other relevant entity) as binary variables representing their activation (ON state) or inactivation (OFF state), along with their interactions inside the cell (A). The interactions between genes determine the behavior of the entire network over time (the dynamics), which is meant to represent the change of cell behavior. The networks can be initialized in any combination of states, which allows the exploration of all possible states. BNs will tend to a type of 'attractor' regardless of their initial configuration, and the collection of states leading to a given attractor is called 'basin of attraction' (B). A single network can reach different attractors, depending on its configuration, and these attractors can be a single state (point attractor) or a collection of states forming a cycle (cyclic attractor) (C). The potential of BNs to model multicellular development stems from the fact that, with a single network, one can reach many stable phenotypes, just as multiple states (cell types) are eventually reached in a developing zygote, which is modeled by crossing the boundaries of a separatrix (B) by some perturbation.

molecules) are updated at the same time, and asynchronous schedules, where a random set of nodes are updated at each time step. A synchronous update produces deterministic dynamics, whereas an asynchronous update generates stochastic dynamics. Other stochastic variants of BNs include a framework that considers propensity parameters for updating each node [51] and another that considers multiple functions for each node that can be selected from a probability distribution at each step (e.g., [52]).

As mentioned earlier, the states and temporal evolution of BNs are discrete. More formally, consider a set $x_1, \ldots, x_n$ of discrete variables, each of which can take values of $\{0, 1\}$. If so, a BN is a function $= f_1, \ldots, f_n : \{0,1\}^n \rightarrow \{0,1\}$, where each coordinate function $f_i : \{0,1\}^n \rightarrow \{0,1\}$ describes how the future value of x_i depends on the present value of the input variables. The dynamics of a synchronous dynamic BN are given by the difference equation $x(t + 1) = F(x(t))$. States that the system stabilizes are called **point attractors** when $F(x) = x$, and **cyclic attractors** when $F^r(x) = x$, for some $r > 1$.

In the BN formalism, the attractors of a system are usually associated with certain phenotypes of the cell, such as proliferation or a differentiated cell type, as originally proposed by Waddington [40], and later formalized by Kauffman [35,36] and Thomas [37]. For example, Mendoza [53] used BNs to model the differentiation process in T helper (Th) cells, recovering attractors representing faithfully the different Th cell types (Th0, Th1, and Th2).

In BN modeling, many features influence the dynamics of the models. For instance, the presence of special patterns in the network, such as feedback loops [54,55] or other types of network motif [56], can indicate the presence of multistability or oscillations. Additionally, there are a variety of

Boolean update rules used to produce the network dynamics. In particular, nested canalizing rules [57] are frequently used, because they show up in published models of regulatory networks and they exhibit more stable dynamics than networks comprising random Boolean rules [57–60].

How Can Boolean Networks Be Used to Model the Evolution of Cell Differentiation?

BNs can make appropriate models to represent the different evolutionary mechanisms leading to cellular differentiation (i.e., those presented in Box 2). For instance, genomic innovation can be easily modeled in a BN by incorporating changes in the set of nodes (the variables $x_1, \ldots, x_n$) or in their update functions. A sensible question to formulate in this way is whether some nodes are more (or less) relevant to innovate the repertoire of cell types, and whether the dynamics of the evolved system(s) are robust, fragile, or antifragile (e.g., [61]). and then the cell types are captured by the attractors of the Boolean system. Likewise, for the diversification of functions via **division of labor**, one could consider special patterns in the wiring diagram (e.g., network motifs [56]) such that the dynamic behavior of a cell is functionally coupled with another, representing an integrated multicellular system. For the integration of a life cycle (i.e., the 'temporal to spatial' transition to multicellularity), one might need to consider multiscale models or hybrid models, which are research topics under active development (although see [62], where the authors use a BN with environmentally controlled thresholds that eventually evolve into multicellular groups). The emergence of differentiation via symmetry breaking is a problem of relating structure with dynamics (i.e., given some initial BN system, will it develop into distinct attractors?) (Figure 3). It is common when analyzing all possible attractors in a BN to explore multiple initial conditions, but in fact one can perturbate (or 'control') the same system and reach multiple attractors within the same basin (e.g., [63]), to cross the **separatrix** boundaries depicted in Figure 3B. Finally, the use of perturbations and stochasticity in the updating rules can also help in modeling the evolution of cellular differentiation via environmental stress or phenotypic plasticity (see [64,65]).

Limitations and Prospects of the Boolean Formalism

For some researchers, it might appear that the lack of rate constants in BNs is a fundamental limitation when matching empirical studies on molecular dynamics (as explained in [66]). However, the classification of cellular behaviors as distinct 'types' is ultimately a qualitative distinction and, thus, BN models appear apt for this task. Perhaps a more relevant limitation of BN is the complexity of the existing algorithms for their analysis, because the number of possible transient states scales exponentially with the number of nodes in a BN [number of possible transient states $= 2^{(\text{n. of nodes})}$]. However, as more data and models are explored, novel methods of BN reduction and analysis are formulated, producing more efficient tools (e.g., [67–69]). Finally, another limitation of BN models is that they have not been widely used to model evolutionary processes at different scales [70], although we have discussed in the previous paragraph some potential ways this can happen and, ultimately, this limitation is more a reflection of current research interests and not fundamental constraints of BN.

Concluding Remarks and Future Perspectives

Multicellularity has evolved independently in diverse phyla (both prokaryotic and eukaryotic) and, in most instances, it has resulted in the appearance of different cell types. Moreover, the number of cell types appears to be correlated with increasing body size (the 'size–complexity rule'; Box 1). A large body of evidence suggests that the unicellular-to-multicellular evolutionary transition involved similar, if not identical, **dynamical patterning modules** (e.g., cell-to-cell adhesion and intercellular communication), but that this transition used different means to achieve them (e.g., carbohydrate versus protein cellular adhesives in land plants and metazoans, respectively).

Outstanding Questions

Are there universal mechanisms underlying the evolution of cell types in all multicellular lineages, or does each lineage have particular mechanisms?

Why do clonal multicellular organisms have a higher number of cell types compared with aggregative multicellularity?

How can we interpret the data generated by single-cell profiling such that the concept of 'cell type' includes a dynamical component?

How can we develop theoretical models that are compatible with data but that also allow us to find new insights into the evolution of cellular differentiation?

How can we develop models that incorporate evolutionary processes, such as novel cell types, while being robust to perturbations?

In this review, we restricted our discussion to some conceptual and theoretical issues necessary for understanding the dynamics and evolution of cellular differentiation. In particular, we focused on how cell-type diversification evolves, and what types of theoretical tools can help disentangle some of the main ideas behind this process. Interpreting the rapidly expanding empirical evidence remains problematic, largely because little effort has gone into contextualizing these results within a theoretical framework. In essence, this work has been hampered by important conceptual limitations, such as the lack of precise and biologically accurate definitions of key concepts (e.g., what is a 'cell type'), and the ambiguities that often result when trying to solve philosophical issues via data analysis.

Perhaps the greatest challenge is that any explanation for cell-type diversification should be canonical, in the sense that it must apply equally to the diverse ways in which cell-type differentiation has evolved (i.e., to all routes to multicellularity, see Outstanding Questions). To meet this challenge, we focused on the utility of a Boolean modeling approach, which shows great promise and versatility in providing deep insights across a spectrum of biological applications, given that it does not rely heavily on specific data or a particular model system. Indeed, one of the great advantages of BNs is that they allow researchers to ask concrete questions while maintaining generality and, as a result, have the potential to provide a formal explanation for how cell types evolve in many contexts.

Despite its potential, however, it is clear that the application of BNs to unravel cell-type differentiation faces persistent challenges. An ever-present caveat to mathematical formalisms is that we cannot know what the relevant level of description for a particular system is (molecular, cellular, multicellular, minutes, hours, or years?). However, cellular differentiation is fundamentally a qualitative change in the behavior of a system (via, for instance, symmetry breaking). This is why BNs represent an encouraging approach, and their use to model cellular differentiation (and its evolution) holds promise. However, one must be cautious with abstract models such as these [26], and they must be continuously and rigorously tested against reliable empirical data. This, in turn, requires the development of technologies that avoid analytical pitfalls, such as false readings of single-cell functional heterogeneity, and the associated analysis pipelines. Therefore, it is crucial that future research develops a dynamic reciprocity between theorists and empiricists, because only then will we be able to break current constraints on our ability to reveal deep insights into the fundamental question of how cells differentiate.

Acknowledgments

P.M-Z., R.P., M.G., A.V.C., D.M., and K.J.N. attended a summit and a follow-up workshop at the Mathematical Bioscience Institute at The Ohio State University (MBI-OSU), supported by the NSF grant DMS-1839810, where this paper was conceived. The authors thank Matthew Thompson and Janet Best for assistance. A.V.C. received support from the Simons Foundation (516088). P.M-Z., R.P., and W.C.R. were supported by NSF grants IOS-1656549 and DEB-1845363. P.M-Z. acknowledges the members of the Ratcliff Lab for engaging discussions on an earlier version of this manuscript.

References

1. Heim, N.A. *et al.* (2017) Hierarchical complexity and the size limits of life. *Proc. R. Soc. B Biol. Sci.* 284, 20171039
2. Szathmáry, E. and Smith, J.M. (1995) The major evolutionary transitions. *Nature* 374, 227–232
3. Bell, G. and Mooers, A.O. (1997) Size and complexity among multicellular organisms. *Biol. J. Linn. Soc.* 60, 345–363
4. Brunet, T. and King, N. (2017) The origin of animal multicellularity and cell differentiation. *Dev. Cell* 43, 124–140
5. Fisher, R.M. *et al.* (2013) Group formation, relatedness, and the evolution of multicellularity. *Curr. Biol.* 23, 1120–1125
6. Wagner, D.E. *et al.* (2018) Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science (80-.)* 360, 981–987
7. Briggs, J.A. *et al.* (2018) The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science (80-.)* 360, eaar5780
8. Svensson, V. *et al.* (2018) Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13, 599–604
9. Michod, R.E. and Roze, D. (2001) Cooperation and conflict in the evolution of multicellularity. *Heredity (Edinb)* 86, 1–7
10. Knoll, A.H. (2011) The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* 39, 217–239
11. Ratcliff, W.C. *et al.* (2012) Experimental evolution of multicellularity. *Proc. Natl. Acad. Sci.* 109, 1595–1600
12. Herron, M.D. *et al.* (2019) De novo origins of multicellularity in response to predation. *Sci. Rep.* 9, 1–9

13. Grosberg, R.K. and Strathmann, R.R. (2007) The evolution of multicellularity: a minor major transition? *Annu. Rev. Ecol. Evol. Syst.* 38, 621–654
14. Tarnita, C.E. *et al.* (2013) Evolutionary construction by staying together and coming together. *J. Theor. Biol.* 320, 10–22
15. Van Gestel, J. and Tarnita, C.E. (2017) On the origin of biological construction, with a focus on multicellularity. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11018–11026
16. Herron, M.D. *et al.* (2013) Cellular differentiation and individuality in the 'minor' multicellular taxa. *Biol. Rev.* 88, 844–861
17. Fisher, R.M. *et al.* (2020) The evolution of multicellular complexity: the role of relatedness and environmental constraints. *Proc. R. Soc. B Biol. Sci.* 287, 20192963
18. Clarke, E. (2011) Plant individuality and multilevel selection theory. In *The Major Transitions Revisited* (Sterelny, K. and Calcott, B., eds), pp. 227–250, MIT Press
19. Kiss, E. *et al.* (2019) Comparative genomics reveals the origin of fungal hyphae and multicellularity. *Nat. Commun.* 10, 1–13
20. Bastiaans, E. *et al.* (2016) Experimental evolution reveals that high relatedness protects multicellular cooperation from cheaters. *Nat. Commun.* 7, 1–10
21. Amado, A. *et al.* (2018) A theoretical approach to the size-complexity rule. *Evolution (N. Y)* 72, 18–29
22. Clevers, H. *et al.* (2017) What is your conceptual definition of 'cell type' in the context of a mature organism? *Cell Syst.* 4, 255–259
23. Wagner, D.E. and Klein, A.M. (2020) Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* 21, 410–427
24. Angerer, P. *et al.* (2017) Single cells make big data: new challenges and opportunities in transcriptomics. *Curr. Opin. Syst. Biol.* 4, 85–91
25. Arendt, D. *et al.* (2016) The origin and evolution of cell types. *Nat. Rev. Genet.* 17, 744
26. Newman, S.A. (2020) Cell differentiation: what have we learned in 50 years? *J. Theor. Biol.* 485, 110031
27. Budnik, B. *et al.* (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* 19, 161
28. Komin, N. and Skupin, A. (2017) How to address cellular heterogeneity by distribution biology. *Curr. Opin. Syst. Biol.* 3, 154–160
29. Symmons, O. and Raj, A. (2016) What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Mol. Cell* 62, 788–802
30. Adler, M. *et al.* (2019) Continuum of gene-expression profiles provides spatial division of labor within a differentiated cell type. *Cell Syst.* 8, 43–52
31. Trapnell, C. *et al.* (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381
32. Pliner, H.A. *et al.* (2019) Supervised classification enables rapid annotation of cell atlases. *Nat. Methods* 16, 983–986
33. Luecken, M.D. and Theis, F.J. (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15, e8746
34. Tyson, J.J. and Novak, B. (2020) A dynamical paradigm for molecular cell biology. *Trends Cell Biol.* 30, 504–515
35. Kauffman, S.A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467
36. Kauffman, S. (1969) Homeostasis and differentiation in random genetic control networks. *Nature* 224, 177–178
37. Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theor. Biol.* 42, 563–585
38. Trigos, A.S. *et al.* (2018) How the evolution of multicellularity set the stage for cancer. *Br. J. Cancer* 118, 145–152
39. Bertolaso, M. and Dieli, A.M. (2017) Cancer and intercellular cooperation. *R. Soc. Open Sci.* 4, 170470
40. Waddington, C.H. (2014) *The Strategy of the Genes*, Routledge
41. Gershenson, C. (2004) Introduction to random Boolean networks. *arXiv* Published online August 2, 2004. https://arxiv.org/abs/nlin/0408006
42. Arias Del Angel, J.A. *et al.* (2018) Cell-fate determination in *Myxococcus xanthus* development: network dynamics and novel predictions. *Develop. Growth Differ.* 60, 121–129
43. Li, F. *et al.* (2004) The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci.* 101, 4781–4786
44. Ríos, O. *et al.* (2015) A Boolean network model of human gonadal sex determination. *Theor. Biol. Med. Model.* 12, 26
45. Espinosa-Soto, C. *et al.* (2004) A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16, 2923–2939
46. Veliz-Cuba, A. and Stigler, B. (2011) Boolean models can explain bistability in the lac operon. *J. Comput. Biol.* 18, 783–794
47. Font-Clos, F. *et al.* (2018) Topography of epithelial–mesenchymal plasticity. *Proc. Natl. Acad. Sci. U. S. A.* 115, 5902–5907
48. Veliz-Cuba, A. *et al.* (2014) Piecewise linear and Boolean models of chemical reaction networks. *Bull. Math. Biol.* 76, 2945–2984
49. Veliz-Cuba, A. *et al.* (2012) On the relationship of steady states of continuous and discrete models arising from biology. *Bull. Math. Biol.* 74, 2779–2792
50. Veliz-Cuba, A. *et al.* (2010) Polynomial algebra of discrete models in systems biology. *Bioinformatics* 26, 1637–1643
51. Murrugarra, D. *et al.* (2012) Modeling stochasticity and variability in gene regulatory networks. *EURASIP J. Bioinforma. Syst. Biol.* 2012, 5
52. Shmulevich, I. and Dougherty, E.R. (2010) *Probabilistic Boolean Networks - The Modeling and Control of Gene Regulatory Networks*, SIAM
53. Mendoza, L. (2006) A network model for the control of the differentiation process in Th cells. *Biosystems* 84, 101–114
54. Remy, E. and Ruet, P. (2008) From minimal signed circuits to the dynamics of Boolean regulatory networks. *Bioinformatics* 24, i220–i226
55. Sontag, E. *et al.* (2008) The effect of negative feedback loops on the dynamics of Boolean networks. *Biophys. J.* 95, 518–526
56. Alon, U. (2019) *An Introduction to Systems Biology: Design Principles of Biological Circuits* (2nd edn), Chapman & Hall/CRC Computational Biology Series
57. Li, Y. *et al.* (2013) Boolean nested canalizing functions: a comprehensive analysis. *Theor. Comput. Sci.* 481, 24–36
58. Kauffman, S. *et al.* (2004) Genetic networks with canalyzing Boolean rules are always stable. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17102–17107
59. Murrugarra, D. and Dimitrova, E.S. (2015) Molecular network control through Boolean canalization. *EURASIP J. Bioinform. Syst. Biol.* 2015, 9
60. Joo, J. Il *et al.* (2018) Determining relative dynamic stability of cell states using Boolean network model. *Sci. Rep.* 8, 1–14
61. Kim, H. *et al.* (2020) Antifragility predicts the robustness and evolvability of biological networks through multi-class classification with a convolutional neural network. *arXiv* Published online February 4, 2020. https://arxiv.org/abs/2002.01571
62. Staps, M. *et al.* (2019) Emergence of diverse life cycles and life histories at the origin of multicellularity. *Nat. Ecol. Evol.* 3, 1197–1205
63. Kim, J. *et al.* (2013) Discovery of a kernel for controlling biomolecular regulatory networks. *Sci. Rep.* 3, 1–9
64. Wagner, G.P. *et al.* (2019) Stress-induced evolutionary innovation: a mechanism for the origin of cell types. *BioEssays* 41, 1800188
65. Schlichting, C.D. (2003) Origins of differentiation via phenotypic plasticity. *Evol. Dev.* 5, 98–105
66. Bornholdt, S. (2008) Boolean network models of cellular regulation: prospects and limitations. *J. R. Soc. Interface* 5
67. Veliz-Cuba, A. (2011) Reduction of Boolean network models. *J. Theor. Biol.* 289, 167–172
68. Murrugarra, D. *et al.* (2016) Identification of control targets in Boolean molecular network models via computational algebra. *BMC Syst. Biol.* 10, 94
69. Zañudo, J.G.T. *et al.* (2017) Structure-based control of complex networks with nonlinear dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 114, 7234–7239
70. Borriello, E. *et al.* (2020) Cell phenotypes as macrostates of the GRN dynamics. *J. Exp. Zool. Part B Mol. Dev. Evol.* 334, 213–224

71. Niklas, K.J. *et al.* (2014) The number of cell types, information content, and the evolution of complex multicellularity. *Acta Soc. Bot. Pol.* 83, 337–347
72. True, J.R. and Carroll, S.B. (2002) Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* 18, 53–80
73. Olson, B.J.S.C. and Nedelcu, A.M. (2016) Co-option during the evolution of multicellular and developmental complexity in the volvocine green algae. *Curr. Opin. Genet. Dev.* 39, 107–115
74. Hanschen, E.R. *et al.* (2016) The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat. Commun.* 7, 1–10
75. Richter, D.J. *et al.* (2018) Gene family innovation, conservation and loss on the animal stem lineage. *eLife* 7, e34226
76. Nedelcu, A.M. and Michod, R.E. (2006) The evolutionary origin of an altruistic gene. *Mol. Biol. Evol.* 23, 1460–1464
77. Featherston, J. *et al.* (2018) The 4-celled *Tetrabaena socialis* nuclear genome reveals the essential components for genetic control of cell number at the origin of multicellularity in the volvocine lineage. *Mol. Biol. Evol.* 35, 855–870
78. Glassford, W.J. *et al.* (2015) Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. *Dev. Cell* 34, 520–531
79. Pires, N.D. and Dolan, L. (2012) Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 508–518
80. De Almeida, A.M.R. *et al.* (2014) Co-option of the polarity gene network shapes filament morphology in angiosperms. *Sci. Rep.* 4, 1–9
81. Rokas, A. (2008) The molecular origins of multicellular transitions. *Curr. Opin. Genet. Dev.* 18, 472–478
82. Oud, B. *et al.* (2013) Genome duplication and mutations in ACE2 cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* 110, E4223–E4231
83. Sebé-Pedrós, A. *et al.* (2017) The origin of Metazoa: a unicellular perspective. *Nat. Rev. Genet.* 18, 498
84. Sebé-Pedrós, A. *et al.* (2018) Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* 2, 1176–1188
85. Sebé-Pedrós, A. *et al.* (2016) The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* 165, 1224–1237
86. Arendt, D. *et al.* (2019) Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* 56, 144–152
87. Rossetti, V. *et al.* (2010) The evolutionary path to terminal differentiation and division of labor in cyanobacteria. *J. Theor. Biol.* 262, 23–34
88. Rueffler, C. *et al.* (2012) Evolution of functional specialization and division of labor. *Proc. Natl. Acad. Sci. U. S. A.* 109, E326–E335
89. Matt, G.Y. and Umen, J.G. (2018) Cell-type transcriptomes of the multicellular green alga *Volvox carteri* yield insights into the evolutionary origins of germ and somatic differentiation programs. *G3 (Bethesda)* 8, 531–550
90. Cooper, G.A. and West, S.A. (2018) Division of labour and the evolution of extreme specialization. *Nat. Ecol. Evol.* 2, 1161–1167
91. Yanni, D. *et al.* (2019) Topological constraints in early multicellularity favor reproductive division of labor. *bioRxiv* Published online November 15, 2019. https://doi.org/10.1101/842849
92. Van Gestel, J. *et al.* (2015) Division of labor in biofilms: the ecology of cell differentiation. *Microb. Biofilms* 3, MB-0002-2014
93. Zhang, Z. *et al.* (2016) Understanding microbial divisions of labor. *Front. Microbiol.* 7, 2070
94. Dayel, M.J. *et al.* (2011) Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* 357, 73–82
95. Mikhailov, K.V. *et al.* (2009) The origin of Metazoa: a transition from temporal to spatial cell differentiation. *Bioessays* 31, 758–768
96. Tecon, R. and Leveau, J.H.J. (2016) Symplasmata are a clonal, conditional, and reversible type of bacterial multicellularity. *Sci. Rep.* 6, 1–10
97. Geerlings, N.M.J. *et al.* (2020) Division of labor and growth during electrical cooperation in multicellular cable bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 117, 5478–5485
98. Black, A.J. *et al.* (2020) Ecological scaffolding and the evolution of individuality. *Nat. Ecol. Evol.* 4, 426–436
99. Chiu, L. and Gilbert, S.F. (2015) The birth of the holobiont: multi-species birthing through mutual scaffolding and niche construction. *Biosemiotics* 8, 191–210
100. Turing, A. (1952) The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 237, 37–72