Spring 2025 – Epigenetics and Systems Biology
Discussion Outline (Systems Biology)
Michael K. Skinner – Biol 476/576
Discussion Session 10:30 am – Noon
Weeks 1 and 2 (January 16, 2025)

Systems Biology

Primary Papers

1. Wu, et al. (2022) Curr Opin Chem Biol 66:102101. (PMID: 34861483)
2. Morelli, et al. (2012) Science 336:187-191. (PMID: 22499940)
3. Gorochowski, et al. (2020) Front. Bioengineering & Biotechnology 8:705. (PMID: 32671054)

Discussion

Student 1 - Ref #1 above
-What omics components are involved in networks?
-What is GWAS and why focus on this?
-How can this approach help medicine?

Student 2 - Ref #2 above
-What are patterning strategies?
-What is mechanical deformation?
-How are gene networks involved?

Student 3 - Ref #3 above
-What is emergence?
-How can synthetic biology be used?
-What are the insights provided in systems biology?

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Network biology bridges the gaps between quantitative genetics and multi-omics to map complex diseases

Si Wu[1], Dijun Chen[2] and Michael P. Snyder[1]

Abstract

With advances in high-throughput sequencing technologies, quantitative genetics approaches have provided insights into genetic basis of many complex diseases. Emerging in-depth multi-omics profiling technologies have created exciting opportunities for systematically investigating intricate interaction networks with different layers of biological molecules underlying disease etiology. Herein, we summarized two main categories of biological networks: evidence-based and statistically inferred. These different types of molecular networks complement each other at both bulk and single-cell levels. We also review three main strategies to incorporate quantitative genetics results with multi-omics data by network analysis: (a) network propagation, (b) functional module-based methods, (c) comparative/dynamic networks. These strategies not only aid in elucidating molecular mechanisms of complex diseases but can guide the search for therapeutic targets.

Addresses

[1] Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[2] State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, 210023, China

Corresponding author: Snyder, Michael P. (mpsnyder@stanford.edu)

Current Opinion in Chemical Biology 2022, **66**:102101

This review comes from a themed issue on **Omics (2022)**

Edited by **Mohan Babu**

For complete overview of the section, please refer to the article collection Omics (2022)

Available online 30 November 2021

https://doi.org/10.1016/j.cbpa.2021.102101

1367-5931/© 2021 Elsevier Ltd. All rights reserved.

Keywords

Quantitative genetics, Multi-omics, Network analysis, Complex diseases.

Introduction

Unlike single-gene diseases, most diseases are complex, that is, caused by a combination of several genetic alterations [1]. The genetics of complex diseases are difficult to study for several reasons. First, the individual effect of each mutation might be small and thus hard to identify [2]. Second, complex diseases often exhibit considerable variation of disease phenotypes [3]. Third, external factors such as environment, diet, and lifestyle interplay with genetic factors, which further complicates their study [4].

The Human Genome Project, completed in 2003 [5], provided new avenues for defining the genetic architecture of human diseases. Recent advances in high-throughput sequencing in addition to early studies using array-based single nucleotide variant (SNV) mapping methods have fueled exceptional growth in quantitative genetics [6]. Quantitative trait locus (QTL) linkage mapping and genome-wide association studies (GWAS) have discovered associations between genetic variations and complex disease traits [7,8]. To complement GWAS that are mostly restricted to one or a small set of diseases, phenome-wide association studies (PheWAS) have been emerged to identify novel genetic associations across a broad spectrum of phenotypes [9,10]. However, the resolution of many QTL mapping studies remains poor, rarely leading to identification of single causal genes [11], and the amount of heritable variation explained by GWAS is often surprisingly low [12]. Intricate interactions between genetic variants and gene products, and gene–environment interaction have been suggested to explain the so-called "missing heritability" [13]. Nonetheless, these caveats make it difficult to translate results into meaningful clinical applications [14]. Additional approaches for translating genetic mapping information into biological discovery are therefore needed.

Classical reductionism, where single causal genes are used to explain disease etiology, is being increasingly challenged, especially with a growing body of evidence indicating the importance of searching beyond the genome, towards interactions between components—the interactome [15]. Research studies are increasingly generating and analyzing multiple omics datasets [16], offering better opportunities to elucidate causative changes that lead to disease or to treatment targets. This development accelerates the emergence of a new paradigm—network biology—combining graph theory, systems biology, and statistical analysis [1]. Advances in graph theory (BOX 1) enable the

Box 1. Basic network properties in graph theory.

Directed vs undirected networks: If the edge connecting two vertices indicates a two-way relationship and can be traversed in both directions, it is in an undirected network. Examples of such networks include PPI networks and gene co-expression networks. If connections between vertices are directional, it is a directed network such as gene regulatory networks, where a gene expression is regulated by a given transcription factor, and the direction usually goes from the transcription factor to the gene. Directed networks are usually more informative than undirected networks.

Degree and degree distribution: The number of edges that one node has is called degree. In random networks, most nodes have a similar number of edges, and their degree distribution follows the Poisson distribution. In contrast, many real-world networks, including most biological networks, are scale-free. This means that their degree distribution follows a power law, as most of the nodes have few links and only a few nodes are densely connected.

Hubs: Hub nodes have the number of links that greatly exceeds the average in a network. In biological networks, hubs tend to be located at the functional center of the interactome, and they are essential genes expressed in multiple tissues. Nevertheless, not all essential genes are disease genes, and disease genes are usually located at the functional periphery of the interactome and tend to be tissue specific.

Betweenness: The extent to which a node participates in the shortest paths connecting other nodes. Nodes with high betweenness, known as "bottlenecks", can be extremely influential in a network in the sense that they are located in critical junctions between hubs and can therefore represent bridges that allow groups of nodes to cross talk to each other.

Closeness: A measure of the average length of the shortest paths from one node to other nodes, which indicates important nodes that can communicate quickly with other nodes of the network.

Network efficiency: The average inverse shortest path length over all pairs of nodes. It quantifies the efficiency of information exchange across the whole network.

Clustering coefficient: Averaging the local clustering-coefficients of all nodes. It is a measure of the degree to which nodes in a network tend to cluster together.

Community: A group of nodes that are more densely connected internally than with the rest of the network, and usually represent the functionally similar molecules in biological networks. A wide array of community detection algorithms has been developed such as fast greedy, matrix–eigenvector-based method, edge-betweenness-based, multi-level modularity optimization algorithm.

accurate prediction of novel disease markers based on their topological properties [17] and the usage of topographic measures such as network density and clustering coefficient as biomarkers of disease state [18]. Since complex diseases usually develop gradually over time, longitudinal multi-omics data with coordinated sets of all layers can provide more comprehensive pathologic landscapes.

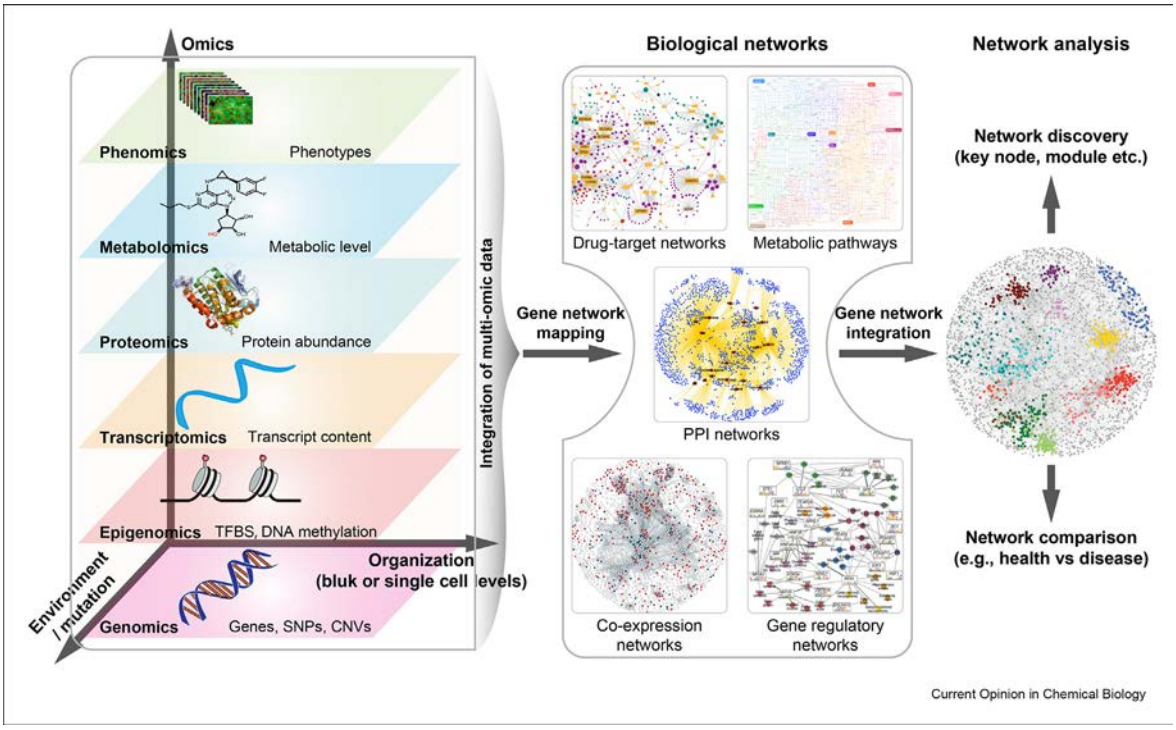
Types of biological networks

Biological networks can either be evidence-based or statistically inferred [19] (Figure 1). The former rely on experimental data aggregated in public databases and/or derived from high throughput technologies; the latter are constructed based on interactions calculated using statistics.

Evidence-based molecular networks

protein–protein interaction (PPI) networks, metabolic networks, regulatory networks, and drug–target networks fall into this category. In PPI networks, nodes are gene-coding proteins connected to each other by physical binding interactions [20]. Large experimental projects, including, but not limited to, yeast two-hybrid assays, as well as immunoprecipitation and/or tandem affinity purification coupled to mass spectrometry, have been undertaken in recent years, yielding approximately 53,000 binary human PPIs [21]. Mosca et al. [22] proposed Interactome 3D, a resource for 3D structural annotation at atomic resolution for over 12,000 PPIs in eight model organisms. Later the same group presented dSysMap, a tool to map disease-related mutations onto 3D PPI interactomes [23]. Both studies have advanced the PPI filed, filling in the gap between systemic and reductionist approaches. Metabolic networks are collections of biochemical reactions between metabolites as well as between metabolites and enzymes, with KEGG (https://www.genome.jp/kegg/) and BiGG (http://bigg.ucsd.edu/) the two most widely used databases. Regulatory networks contain directed links that represent how one gene activates other genes, resulting in more accurate inferences on causal relationships between transcription factors and genes or kinases and their substrates (post-translational modifications) [24]. Important regulatory databases include the ChIP-chip- and ChIP-seq- derived ENCODE (https://www.encodeproject.org/), RegulomeDB (https://

Figure 1



Comprehensive network analysis based on multi-omics data. Multi-omics data (genomics, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics) are collected. Multi-omics data enable us to construct different types of biological networks (PPI networks, gene regulatory networks, metabolic networks, drug–target networks, co-expression networks, and more) at bulk and single cell levels. The complex interactions in these networks not only reflect genetic alterations, but also external factors such as environment, diet, lifestyle, drug administration, etc. These factors interact with each other and can be integrated for further downstream network analysis (e.g., network propagation, discovery of key nodes and functional modules, network comparisons) to generate a comprehensive molecular landscape of complex diseases, as well as to evaluate drug effects.

regulomedb.org/regulome-search/), UniPROBE (http://thebrain.bwh.harvard.edu/uniprobe/) and JASPAR (http://jaspar.genereg.net/) databases, protein–DNA interaction databases such as TRANSFAC (https://genexplain.com/transfac/), EdgeExpressDB (https://fantom.gsc.riken.jp/), MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/), and B-cell interactome (http://califano.c2b2.columbia.edu/b-cell-interactome), as well as databases for post-translational modifications (PhosphoELM (http://phospho.elm.eu.org/), PhosphoSite (https://www.phosphosite.org), PHOSIDA (http://phosida/index.aspx), NetPhorest (http://www.netphorest.info/index.shtml), and CBS (http://www.cbs.dtu.dk/services/)). With many of the databases still incomplete, the regulatory networks are themselves immature [25]. Lastly, drug–target networks [26] are mainly constructed based on public databases such as DrugBank (https://go.drugbank.com/) and SNAP (https://snap.stanford.edu/). These different interactions can be directly assembled into biological networks. A major limitation of evidence-based networks is their reliance on public databases, which remain incomplete and noisy (i.e., false positives). This has led to the development of many statistical inference approaches.

Statistically inferred networks

These networks include gene co-expression networks and genetic networks. Co-expression networks are based on the "guilt-by-association" principle, whereby genes with similar expression profiles likely have similar functions. Genetic networks are built to investigate how the impact of one dysregulated gene spreads along the links throughout the entire molecular network to influence specific disease phenotypes [27]. Some studies have used different molecular layers to construct heterogeneous networks, such as gene–metabolite networks [28] and protein–metabolite networks [29]. There are three methods to calculate the metric used to build edges in statistically inferred networks [30]: (a) correlation-based methods construct an adjacency matrix by calculating the correlations (Pearson or Spearman correlations) or the partial correlations between pairs of molecules. These methods are primarily used to obtain linear correlations among variables; (b) mutual information-based methods measure the mutual dependence between each pair of molecules based on their profiles. These methods can capture both linear and non-linear relationships but can be computationally

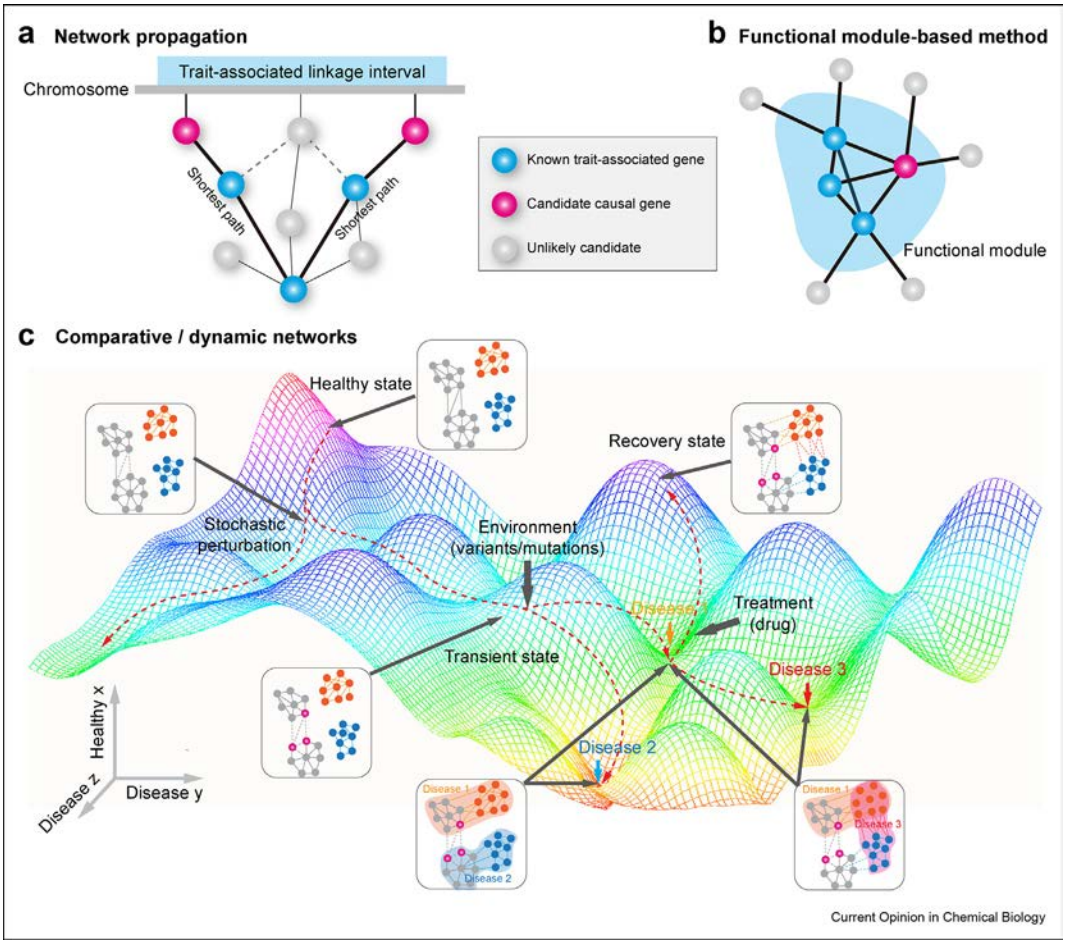
expensive; () tree-based methods decompose the construction of a regulatory network with p genes into p different regression problems. In each of the regression, the expression pattern of the target gene is predicted from the expression patterns of all the other genes (input genes) using tree-based ensemble methods. The importance of an input gene in the prediction of the target gene expression pattern is taken as an indication of a putative regulatory link. These methods are non-parametric and can be applied to high-dimensionality data, demonstrating high accuracy [19].

These different types of networks complement each other and provide distinct insights into cellular interaction mechanisms (Figure 1). They are mainly based on bulk omics data, which are typically limited to the understanding at the organism level [31]. The recent emergence of single-cell multi-omics technologies is able to accelerate the investigation of cellular heterogeneity [32], allowing the construction of cell-type-specific molecular networks associated with certain complex diseases. Single-cell omics networks include GCN (gene co-expression networks), GRN (gene regulatory networks), MGRN (methylation-associated gene regulatory networks), and CGN (*cis*-regulatory gene interaction networks) [33]. The integration of these networks has been applied in complex diseases such as leukemia [34], breast cancer [35], and melanoma [36]. Comparing single-cell omics networks between different cell populations allows us to identify cell-developmental trajectories that may reflect cell lineages, disease progression, and to explore rare cell populations potentially associated with complex diseases [37].

Strategies to bridge over quantitative genetics and multi-omics

How can we incorporate the obtained associations between phenotypes and genotypes from quantitative genetics results to the two types of biological networks? The following three general strategies provide several insightful examples (Figure 2).

Figure 2



Three main network analysis strategies are proposed to bridge over quantitative genetics and multi-omics. (**a**) Network propagation; (**b**) Functional module-based method; (**c**) Comparative/dynamic networks.

Strategy 1: use QTL candidate genes as seed to propagate

If GWAS loci or QTL genes for diseases are available, we can use them as anchors for inferring causality. One straightforward way is to search for neighbor genes directly connected to the causal genes in the interaction networks [38]. However, this approach may bring in a high number of suggestive candidate interactions, especially in large biological networks, making follow-up indiscriminate experimental validation prohibitive.

Another implementation of this strategy is to combine both genotypic results (causal genes from quantitative genetics approaches) with phenotypic data (omics data showing disease effects) by leveraging network topological properties. Genotypic variation is considered the source of perturbations, and molecules with phenotypic changes are considered as the targets of the perturbations. The shortest path connecting causal genes with targets is often thought to infer causality [39]. The intermediate vertices on this shortest path are likely the key components of the perturbed pathways. Several studies have applied this method to identify the affected pathways/modules from seed genes to their targets [40,41]. If multiple causal genes act as seeds, then the Steiner tree-based method is widely used, where key elements are identified by minimizing the total edge cost [42]. Bechet et al. employed this method to successfully predict the functional role of an uncharacterized protein COS8 in sphingolipid biosynthesis and TOR signaling [43]. Tuncbag et al. [44] expanded this application to the Steiner Forest and developed a web-based tool "SteinerNet" to discover hidden components in dysregulated pathways by integrating omics data. However, shortest path-based methods have several drawbacks [45]: (a) they do not incorporate other molecular layer data, such as gene-expression data; (b) they assume that the shortest path contains the most informative pathways with the key components, which is not always true; (c) they ignore the possibility of multiple shortest paths from seed genes to target genes, especially in large and complicated networks.

To address these limitations, flow-based methods for finding the most possible dysregulated paths were developed [45], where the current in an electronic circuit is mimicked and the fraction of flow—the probability of the biological information propagates through a given node or edge—is calculated based on probabilistic theory. This method is superior for incorporating additional data to retrieve information-propagation pathways with more confidence. Kim et al. [46] applied this method and incorporated it with eQTL, successfully narrowing down the causal genes and underlying dysregulated pathways, and identifying several hub nodes that are well-documented as important regulators in glioma. This strategy largely alleviates the common issue in quantitative genetics approaches that mapped loci may still contain dozens of candidate causal genes with false positives.

Strategy 2: map QTL candidate genes into functional modules

The causal genes from QTL mapping can be placed into a molecular network, and then searched for network modules enriched with the altered genes. This method can increase statistical power to identify a subset of genes in the enriched modules that may not show significant signals in either GWAS or QTL mapping, but still play important roles within modules by mediating interactions between causal disease genes and other associated components. This approach can partly overcome the main shortcomings of quantitative genetics that rare germline variations are difficult to discover and distinguish from noises due to their rarity. The module-based strategy may identify this accumulation effect. Several benchmark studies have successfully integrated genetics results with network modeling, and thereby identified key modules that contribute to diseases [47–49]. For instance, Marbach et al. [50] collected genomics, epigenomics, and transcriptomics data and constructed 394 cell-type- and tissue-specific gene regulatory networks. They then overlaid the GWAS results, including variants that do not pass GWAS significance, onto these gene regulatory networks, and identified cell-type- or tissue-specific modules that are enriched for genetic variants in human diseases.

Strategy 3: comparative network analysis

Biological networks undergo dynamic rewiring in response to different disease states, disease development stages, environmental stresses, drug treatment, or through evolutionary time [51]. There are two main branches in the comparative network analysis strategy. The first focuses on conserved modules over the time or different status, underlying preserved core functions [51]. This approach has been widely applied in evolution analysis across different species for prediction of novel gene function [52] and mechanisms of drug action [53]. A different branch focuses on differential network structures by comparing networks across different conditions. Once the different networks are constructed, one can search for loss and gain of correlations with the seed genes (e.g., QTL causal genes) in different disease states to identify dysregulated genes. Another way is to compare the topological structure (BOX1) in different networks, as the topological rewiring can trigger changes in specific disease-associated modules. In one interesting study, gene co-expression networks of normal and prostate-cancer samples were compared. The authors detected prostate-cancer-specific modules involved in RAD50 and telomeric repeat-binding factor 2 [54]. Comparative network approach can be carried out on physical networks

(e.g., PPI, protein−DNA networks) [55], as well as functional networks (e.g., co-expression networks, metabolic networks) [56], which can shed light on disease progression mechanisms and drug responses.

Leveraging the advantages of network analysis, we can integrate quantitative genetics results with rich multi-omics resources to potentially address many important questions in complex diseases. The four broad categories of these questions that may be answered by the three strategies are summarized in Supplementary Table 1.

Conclusion

Network biology provides an ideal tool to link quantitative genetics and multi-omics data and will likely play a major role in disease-biomarker identification and novel-therapy discovery in the next decade. This paradigm is highly tailored and suitable for complex diseases with highly heterogeneous molecular basis and ones that can be influenced by intricate factors (e.g., environment, diet, drug administration) simultaneously. The explosive advancements in sequencing and mass spectrometry technologies, as well as the flourishing development of novel network-modeling algorithms, have paved ways for the next generation of precision medicine. That said, network approaches are not free of limitations, which include incomplete interactome databases due to inherent detection limitations of high-throughput experiments, lack of ground truth for the validation of constructed molecular networks, and their common focus on static conditions, thereby overlooking the dynamic changes of molecular interactions [51]. Despite these limitations, network biology, a revolutionary approach, bridges over quantitative genetics approaches and multi-omics technologies, encompassing multiple data types, algorithms, as well as conditions and timepoints at growing scale and depth. The integration of single-cell multi-omics networks provides more opportunities to demystify molecular mechanisms of complex diseases in individual cells, as well as guide clinical therapeutical strategy in personalized medicine.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
M.P.S. is cofounder and scientific advisor of Personalis, Qbio, SensOmics, January AI, Mirvie, Protos, NiMo, Onza and is on the advisory board of Genapsys.

Acknowledgements

This work is supported by the NIH grants NIH 5RM1HG00773508 and NIH 1U2CCA233311-01. We thank Dr. Hezi Tenenboim for his kind editorial assistance and Dr. Yariv Brotman for the helpful discussion. Dijun Chen appreciates the support from the Nanjing University Deng Feng Scholars Program and the National Natural Science Foundation of China (32070656).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cbpa.2021.102101.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Hu JX, Thomas CE, Brunak S: **Network biology concepts in complex disease comorbidities**. *Nat Rev Genet* 2016, **17**: 615−629.

2. Karczewski KJ, Snyder MP: **Integrative omics for health and disease**. *Nat Rev Genet* 2018, **19**:299−310.

3. Turajlic S, Sottoriva A, Graham T, Swanton C: **Resolving genetic heterogeneity in cancer**. *Nat Rev Genet* 2019, **20**:404−416.

4. Li J, Li X, Zhang S, Snyder M: **Gene-environment interaction in the era of precision medicine**. *Cell* 2019, **177**:38−44.

5. Collins FS, Morgan M, Patrinos A: **The human genome project: lessons from large-scale biology**. *Science* 2003, **300**:286.

6. Jakobson CM, Jarosz DF: **What has a century of quantitative genetics taught us about nature's genetic tool kit?** *Annu Rev Genet* 2020, **54**:439−464.

7. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yan Y, Kundu K, Ecker S, *et al.*: **Genetic drivers of epigenetic and transcriptional variation in human immune cells**. *Cell* 2016, **167**:1398−1414.e24.

8. * Investigators kConFab, , ABCTB Investigators, EMBRACE Study, GEMO Study Collaborators, Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, *et al.*: **Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses**. *Nat Genet* 2020, **52**:572−581.
One of the largest GWAS for breast cancer. The authors applied both standard and novel methodologies that account for tumor heterogeneity to identify 32 novel breast cancer susceptibility loci and five of them contained cell-specific enhancers.

9. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, *et al.*: **Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data**. *Nat Biotechnol* 2013, **31**: 1102−1111.

10. Ma J, Yu MK, Fong S, Ono K, Sage E, Demchak B, Sharan R, Ideker T: **Using deep learning to model the hierarchical structure and function of a cell**. *Nat Methods* 2018, **15**: 290−298.

11. Jamann TM, Balint-Kurti PJ, Holland JB: **QTL mapping using high-throughput sequencing**. In *Plant functional genomics: methods and protocols*. Edited by Alonso JM, Stepanova AN, New York: Springer; 2015:257−285.

12. Dai J, Shen W, Wen W, Chang J, Wang T, Chen H, Jin G, Ma H, Wu C, Li L, *et al.*: **Estimation of heritability for nine common cancers using data from genome-wide association studies in Chinese population: estimation of heritability for nine common cancers**. *Int J Cancer* 2017, **140**: 329−336.

13. Génin E: **Missing heritability of complex diseases: case solved?** *Hum Genet* 2020, **139**:103−113.

14. Cano-Gamez E, Trynka G: **From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases**. *Front Genet* 2020, **11**:424.

15. Caldera M, Buphamalai P, Müller F, Menche J: **Interactome-based approaches to human disease**. *Curr Opin Struct Biol* 2017, **3**:88−94.

16. Joshi A, Rienks M, Theofilatos K, Mayr M: **Systems biology in cardiovascular disease: a multiomics approach**. *Nat Rev Cardiol* 2021, **18**:313–330.

17. Yadav A, Vidal M, Luck K: **Precision medicine — networks to the rescue**. *Curr Opin Biotechnol* 2020, **63**:177–189.

18. Yu X, Lai S, Chen H, Chen M: **Protein–protein interaction network with machine learning models and multiomics data reveal potential neurodegenerative disease-related proteins**. *Hum Mol Genet* 2020, **29**:1378–1387.

19. Oulas A, Minadakis G, Zachariou M, Sokratous K, Bourdakou MM, Spyrou GM: **Systems Bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches**. *Briefings Bioinf* 2019, **20**:806–824.

20. Silverbush D, Sharan R: **A systematic approach to orient the human protein–protein interaction network**. *Nat Commun* 2019, **10**:3015.

21. * Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, *et al.*: **A reference map of the human binary protein interactome**. *Nature* 2020, **580**:402–408.
The largest available human protein–protein interaction map so far with approximately 53,000 binary protein interactions. It represents an unprecedented contribution for advancing our understanding of cellular physiological and pathological functions.

22. Mosca R, Céol A, Aloy P: **Interactome3D: adding structural details to protein networks**. *Nat Methods* 2013, **10**:47–53.

23. Mosca R, Tenorio-Laranga J, Olivella R, Alcalde V, Céol A, Soler-López M, Aloy P: **dSysMap: exploring the edgetic role of disease mutations**. *Nat Methods* 2015, **12**:167–168.

24. * Canchi S, Raao B, Masliah D, Rosenthal SB, Sasik R, Fisch KM, De Jager PL, Bennett DA, Rissman RA: **Integrating gene and protein expression reveals perturbed functional networks in Alzheimer's disease**. *Cell Rep* 2019, **28**:1103–1116.e4.
A novel network analysis framework by integrating gene expression profiles with brain tissue-specific protein interaction networks was proposed to identify the four functional clusters with distinct biological processes and the key transcriptional regulators in Alzheimer's disease.

25. Sonawane AR, Weiss ST, Glass K, Sharma A: **Network medicine in the age of biomedical big data**. *Front Genet* 2019, **10**:294.

26. Yõldõrõm MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M: **Drug—target network**. *Nat Biotechnol* 2007, **25**:1119–1126.

27. Costanzo M, Kuzmin E, van Leeuwen J, Mair B, Moffat J, Boone C, Andrews B: **Global genetic networks and the genotype-to-phenotype relationship**. *Cell* 2019, **177**:85–100.

28. Mi K, Jiang Y, Chen J, Lv D, Qian Z, Sun H, Shang D: **Construction and analysis of human diseases and metabolites network**. *Front Bioeng Biotechnol* 2020, **8**:398.

29. Contrepois K, Wu S, Moneghetti KJ, Hornburg D, Ahadi S, Tsai M-S, Metwally AA, Wei E, Lee-McMullen B, Quijada JV, *et al.*: **Molecular choreography of acute exercise**. *Cell* 2020, **181**:1112–1130. e16.

30. Bourdakou MM, Athanasiadis EI, Spyrou GM: **Discovering gene re-ranking efficiency and conserved gene-gene relationships derived from gene co-expression network analysis on breast cancer data**. *Sci Rep* 2016, **6**:20518.

31. Hasin Y, Seldin M, Lusis A: **Multi-omics approaches to disease**. *Genome Biol* 2017, **18**:83.

32. Perkel JM: **Single-cell analysis enters the multiomics age**. *Nature* 2021, **595**:614–616.

33. * Li Y, Ma A, Mathé EA, Li L, Liu B, Ma Q: **Elucidation of biological networks across complex diseases using single-cell omics**. *Trends Genet* 2020, **36**:951–966.
A systematic review about using the emerging single-cell omics data to construct different types of biological networks for advancing our understanding of complex diseases. An interesting concept of "heterogeneous regulatory landscape" is first proposed in this review to demonstrate cell heterogeneity.

34. ** Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, Parks B, Gars E, Liedtke M, Zheng GXY, *et al.*: **Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia**. *Nat Biotechnol* 2019, **37**:1458–1465.
A representative study integrating single-cell multi-omics data to establish a normal epigenetic baseline for healthy blood development, then using it to deconvolve abnormal molecular features in acute leukemia patients. The identified candidate transcription factors can potentially serve as novel drug targets for acute leukemia.

35. Gkountela S, Castro-Giner F, Szczerba BM, Vetter M, Landin J, Scherrer R, Krol I, Scheidmann MC, Beisel C, Stirnimann CU, *et al.*: **Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding**. *Cell* 2019, **176**:98–112. e14.

36. Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, Karras P, Brown D, Chang YH, Debiec-Rychter M, *et al.*: **Toward minimal residual disease-directed therapy in melanoma**. *Cell* 2018, **174**:843–855. e19.

37. ** Iacono G, Massoni-Badosa R, Heyn H: **Single-cell transcriptomics unveils gene regulatory network plasticity**. *Genome Biol* 2019, **20**:110.
Single-cell RNA seq derived regulatory networks from organs and different perturbed systems were constructed and compared for topological properties from graph theory. Key players in organ function and drivers of diseases were identified.

38. Chan SY, Loscalzo J: **The emerging paradigm of network medicine in the study of human disease**. *Circ Res* 2012, **111**:359–374.

39. Cowen L, Ideker T, Raphael BJ, Sharan R: **Network propagation: a universal amplifier of genetic associations**. *Nat Rev Genet* 2017, **18**:551–562.

40. Shih Y-K, Parthasarathy S: **A single source k-shortest paths algorithm to infer regulatory pathways in a gene network**. *Bioinformatics* 2012, **28**:i49–i58.

41. ** Wang R, Simoneau CR, Kulsuptrakul J, Bouhaddou M, Travisano KA, Hayashi JM, Carlson-Stevermer J, Zengel JR, Richards CM, Fozouni P, *et al.*: **Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses**. *Cell* 2021, **184**:106–119.e14.
This study applied an integrative network propagation strategy to identify subnetworks and pathways that were common and distinct in the three coronavirus screens to provide biological insights.

42. Sun Y, Ma C, Halgamuge S: **The node-weighted Steiner tree approach to identify elements of cancer-related signaling pathways**. *BMC Bioinf* 2017, **18**:551.

43. Bailly-Bechet M, Borgs C, Braunstein A, Chayes J, Dagkessamanskaia A, Francois J-M, Zecchina R: **Finding undetected protein associations in cell signaling by belief propagation**. *Proc Natl Acad Sci Unit States Am* 2011, **108**:882–887.

44. Tuncbag N, McCallum S, Huang S -sC, Fraenkel E, SteinerNet: **A web server for integrating "omic" data to discover hidden components of response pathways**. *Nucleic Acids Res* 2012, **40**:W505–W509.

45. Cho D-Y, Kim Y-A, Przytycka TM: **Chapter 5: network biology approach to complex diseases**. *PLoS Comput Biol* 2012, **8**:e1002820.

46. Kim Y-A, Wuchty S, Przytycka TM: **Identifying causal genes and dysregulated pathways in complex diseases**. *PLoS Comput Biol* 2011, **7**:e1001095.

47. Li J, Shi M, Ma Z, Zhao S, Euskirchen G, Ziskin J, Urban A, Hallmayer J, Snyder M: **Integrated systems analysis reveals a molecular network underlying autism spectrum disorders**. *Mol Syst Biol* 2014, **10**:774.

48. Calabrese GM, Mesner LD, Stains JP, Tommasini SM, Horowitz MC, Rosen CJ, Farber CR: **Integrating GWAS and Co-expression network data identifies bone mineral density genes SPTBN1 and MARK3 and an osteoblast functional module**. *Cell Systems* 2017, **4**:46–59. e4.

49. ** Yao V, Kaletsky R, Keyes W, Mor DE, Wong AK, Sohrabi S, Murphy CT, Troyanskaya OG: **An integrative tissue-network approach to identify and test human disease genes**. *Nat Biotechnol* 2018, **36**:1091–1099.

A novel framework called "diseaseQUEST" by combining both computational and experimental efforts was proposed to predict candidate genes for human diseases. The main parts that constitute the computational section are the incorporation of tissue-specific functional networks with quantitative genetics data.

50. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S: **Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases**. *Nat Methods* 2016, **13**:366–370.

51. Mitra K, Carvunis A-R, Ramesh SK, Ideker T: **Integrative approaches for finding modular structure in biological networks**. *Nat Rev Genet* 2013, **14**:719–732.

52. International Frontotemporal Dementia Genomics Consortium, Swarup V, Hinz FI, Rexach JE, Noguchi K, Toyoshiba H, Oda A, Hirai K, Sarkar A, Seyfried NT, *et al.*: **Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia**. *Nat Med* 2019, **25**:152–164.

53. Kapitzky L, Beltrao P, Berens TJ, Gassner N, Zhou C, Wüster A, Wu J, Babu MM, Elledge SJ, Toczyski D, *et al.*: **Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action**. *Mol Syst Biol* 2010, **6**:451.

54. Altay G, Asim M, Markowetz F, Neal DE: **Differential C3NET reveals disease networks of direct physical interactions**. *BMC Bioinf* 2011, **12**:296.

55. Tomkins JE, Dihanich S, Beilina A, Ferrari R, Ilacqua N, Cookson MR, Lewis PA, Manzoni C: **Comparative protein interaction network analysis identifies shared and distinct functions for the human ROCO proteins**. *Proteomics* 2018, **18**:1700444.

56. ** Assi SA, Imperato MR, Coleman DJL, Pickin A, Potluri S, Ptasinska A, Chin PS, Blair H, Cauchy P, James SR, *et al.*: **Subtype-specific regulatory network rewiring in acute myeloid leukemia**. *Nat Genet* 2019, **51**:151–162.

A benchmark paper constructed and compared acute myeloid leukemia (AML) subtype-specific transcriptional networks due to different mutant regulators. The dynamic network rewiring provides insight into the unique sets of genes required for subtype AML development.

the mRNA X and mRNA Y corresponding to protein X and protein Y, respectively. Although protein X and protein Y are coordinated for all four motifs in Fig. 3, this is not the case for their mRNA levels. This can be explained by the disparate time scales of mRNA and protein. Fast-degrading mRNA may exhibit fluctuations with a broad frequency bandwidth. Conversely, slow degradation of proteins filters out fast fluctuations but keeps slow fluctuations. Constitutively expressed mRNA X has both fast and slow fluctuations, but protein X only transmits the slow fluctuations downstream. The result is that the dynamics of mRNA X and mRNA Y are dominated by uncorrelated fast fluctuations, which overshadow their correlated slow fluctuations. On the other hand, protein X and protein Y only contain the better-correlated slow fluctuations. That is, two mRNA species can be mostly uncorrelated with one another, yet produce protein in a coordinated fashion. Gandhi *et al.* (*18*) observed such a circumstance in budding yeast, when they found very little correlation between pairs of transcripts that encode coordinated proteins of the same protein complex, including proteasome and RNA polymerase II subunits. They even found correlation lacking in two alleles of the same gene. In a related study, Taniguchi *et al.* (*27*) analyzed more than 1000 genes in *E. coli* and measured both mRNA and protein copy numbers in single cells. They found that for most genes, even the numbers of mRNA and protein molecules were uncorrelated. These studies suggest that understanding of regulatory phenomena requires one to consider regulation at both the mRNA and the protein level.

From these studies, it is now clear that variability in single-cell measurements contains a wealth of information that can reveal new insights into the regulatory phenomena of specific genes and the dynamic interplay of entire gene networks. As modern imaging techniques begin to beat the diffraction limitations of light (*28*) and flow cytometers become affordable for nearly any laboratory bench (*29*), we find ourselves in the midst of an explosion in single-cell research. With the advent of single-cell sequencing (*30*, *31*), it might be possible to determine the full transcriptome of many single cells in the near future and to determine the full expression distributions and correlations for all genes in the genome. We expect that the approaches described in this review, which have been pioneered with the model microbial systems, will be readily applied to mammalian cells and tissues (*32*, *33*).

References and Notes

1. G.-W. Li, X. S. Xie, *Nature* **475**, 308 (2011).
2. A. Raj, A. van Oudenaarden, *Cell* **135**, 216 (2008).
3. G. Balázsi, A. van Oudenaarden, J. J. Collins, *Cell* **144**, 910 (2011).
4. A. Eldar, M. B. Elowitz, *Nature* **467**, 167 (2010).
5. M. E. Lidstrom, M. C. Konopka, *Nat. Chem. Biol.* **6**, 705 (2010).
6. B. Snijder, L. Pelkmans, *Nat. Rev. Mol. Cell Biol.* **12**, 119 (2011).
7. D. Zenklusen, D. R. Larson, R. H. Singer, *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).
8. M. Thattai, A. van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614 (2001).
9. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, *Science* **280**, 585 (1998).
10. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nat. Methods* **5**, 877 (2008).
11. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, *Cell* **123**, 1025 (2005).
12. J. Peccoud, B. Ycart, *Theor. Popul. Biol.* **48**, 222 (1995).
13. T. B. Kepler, T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
14. J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
15. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, *PLoS Biol.* **4**, e309 (2006).
16. V. Shahrezaei, P. S. Swain, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17256 (2008).
17. S. Iyer-Biswas, F. Hayot, C. Jayaprakash, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **79**, 031911 (2009).
18. S. J. Gandhi, D. Zenklusen, T. Lionnet, R. H. Singer, *Nat. Struct. Mol. Biol.* **18**, 27 (2011).
19. L.-H. So *et al.*, *Nat. Genet.* **43**, 554 (2011).
20. R. Z. Tan, A. van Oudenaarden, *Mol. Syst. Biol.* **6**, 358 (2010).
21. S. L. Bumgarner *et al.*, *Mol. Cell* **45**, 470 (2012).
22. L. M. Octavio, K. Gedeon, N. Maheshri, *PLoS Genet.* **5**, e1000673 (2009).
23. J. M. Pedraza, A. van Oudenaarden, *Science* **307**, 1965 (2005).
24. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, *Science* **307**, 1962 (2005).
25. J. Stewart-Ornstein, J. S. Weissman, H. El-Samad, *Mol. Cell* **45**, 483 (2012).
26. M. J. Dunlop, R. S. Cox III, J. H. Levine, R. M. Murray, M. B. Elowitz, *Nat. Genet.* **40**, 1493 (2008).
27. Y. Taniguchi *et al.*, *Science* **329**, 533 (2010).
28. B. Huang, M. Bates, X. Zhuang, *Annu. Rev. Biochem.* **78**, 993 (2009).
29. L. Bonetta, *Nat. Methods* **2**, 785 (2005).
30. T. Kalisky, P. Blainey, S. R. Quake, *Annu. Rev. Genet.* **45**, 431 (2011).
31. F. Tang *et al.*, *Nat. Methods* **6**, 377 (2009).
32. S. Itzkovitz *et al.*, *Nat. Cell Biol.* **14**, 106 (2012).
33. P. Dalerba *et al.*, *Nat. Biotechnol.* **29**, 1120 (2011).
34. B. Munsky, M. Khammash, *J. Chem. Phys.* **124**, 044104 (2006).
35. D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).

Acknowledgments: This work was funded by the National Science Foundation (ECCS-0835623) and a NIH Pioneer award (1DP1OD003936).

10.1126/science.1216379

Downloaded from http://science.sciencemag.org/ on December 21, 2016

REVIEW

Computational Approaches to Developmental Patterning

Luis G. Morelli,[1,2,3] Koichiro Uriu,[1,4] Saúl Ares,[2,5,6] Andrew C. Oates[1]*

Computational approaches are breaking new ground in understanding how embryos form. Here, we discuss recent studies that couple precise measurements in the embryo with appropriately matched modeling and computational methods to investigate classic embryonic patterning strategies. We include signaling gradients, activator-inhibitor systems, and coupled oscillators, as well as emerging paradigms such as tissue deformation. Parallel progress in theory and experiment will play an increasingly central role in deciphering developmental patterning.

Animal and plant patterns amaze and perplex scientists and lay people alike. But how are the dynamic and beautiful patterns of developing embryos generated? Used appropriately, theoretical techniques can assist in the understanding of developmental processes (*1*–*5*). There is considerable art in this, and the key to success is an open dialogue between experimentalist and theorist. The first step in this dialogue is to formulate a theoretical description of the process of interest that captures the properties and interactions of the most relevant variables of the system at a level of detail that is both useful and tractable. Once formulated, the second step is to analyze the theoretical model. If the model is sufficiently tractable, it may be possible to understand its behavior with "pencil-and-paper" analysis and compare this analytical solution directly with experimental data. Very often, however, the number of variables and the complexity of their interactions preclude this approach, and the behavior of models must be solved or simulated by using computers in order to be understood and compared with data. This combined approach, which we refer to as computational biology, has become popular recently with the availability of powerful computers and increasingly sophisticated numerical algorithms.

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany. [2]Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, 01187 Dresden, Germany. [3]Consejo Nacional de Investigaciones Científicas y Técnicas, Departamento de Física, Universidad de Buenos Aires, Ciudad Universitaria, 1428 Buenos Aires, Argentina. [4]Theoretical Biology Laboratory, RIKEN Advanced Science Institute, Saitama 351-0198, Japan. [5]Logic of Genomic Systems Laboratory, Centro Nacional de Biotecnología–Consejo Superior de Investigaciones Científicas (CSIC), Calle Darwin 3, 28049 Madrid, Spain. [6]Grupo Interdisciplinar de Sistemas Complejos (GISC), Spain.

*To whom correspondence should be addressed. E-mail: oates@mpi-cbg.de

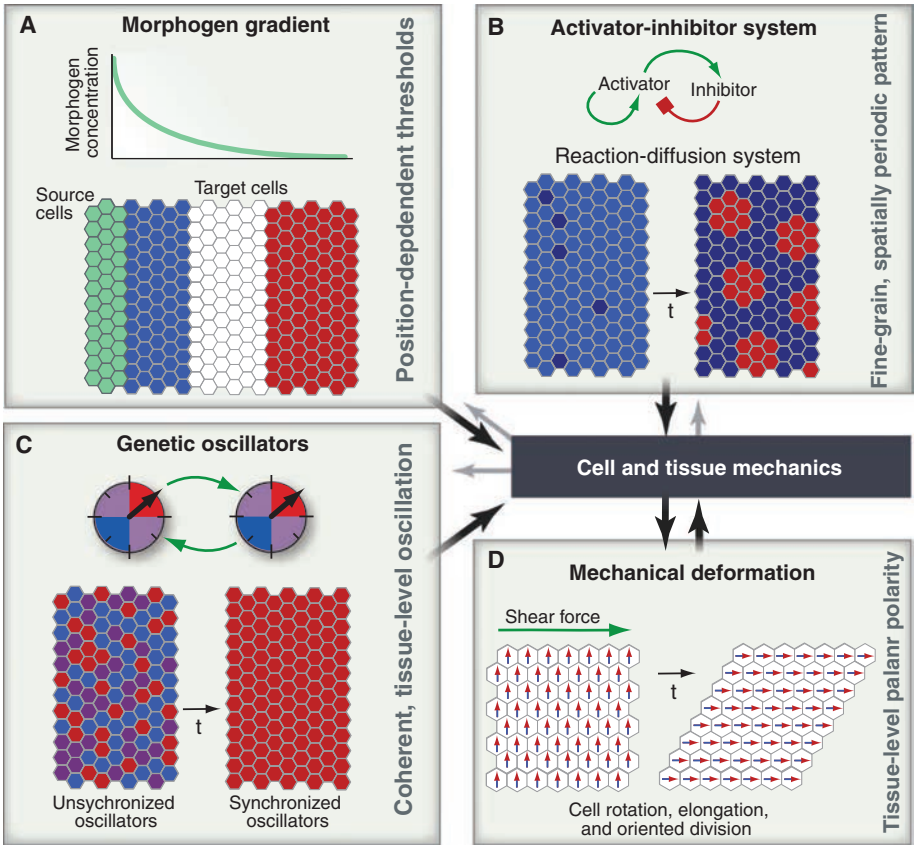


Fig. 1. Patterning strategies. (**A**) Signaling gradients supply global positional information. Horizontal axis is position within target tissue. Morphogen-producing cells are green; cells in tissue take identities (blue, white, and red) according to morphogen concentration. (**B**) Activator-inhibitor systems incorporate local positive and negative feedbacks to generate pattern. Distinct cell types are in red and blue. (**C**) Synchronization of genetic oscillators allows a tissue to generate a coherent temporal rhythm for patterning. In these snapshots, the phase of each oscillating cell is given by its color, which changes over time. (**D**) Tissue deformation can drive patterning reactions. Downstream of patterning information, the dynamic physical properties of tissues drive the morphogenesis of the embryo. *t*, time.

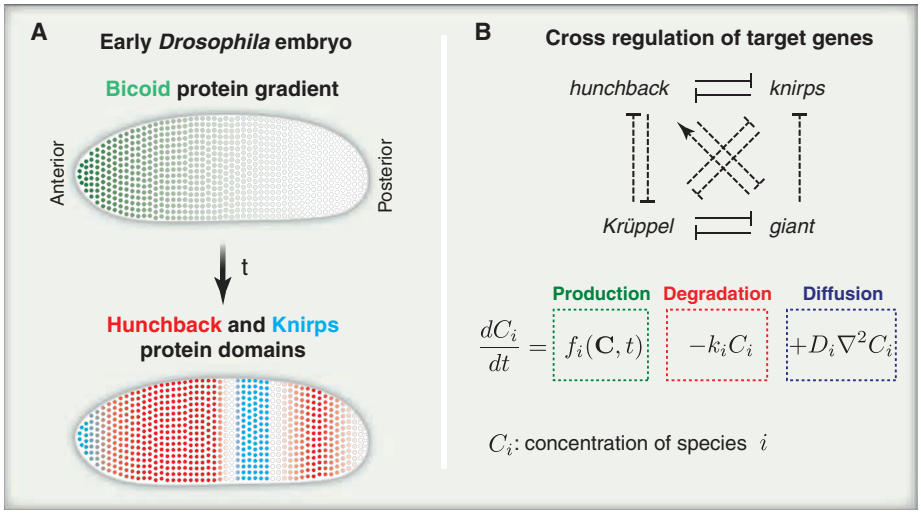


Fig. 2. Patterning with signaling gradients. (**A**) Schematic of early fruit fly embryo showing the maternal gradient of Bicoid protein at cycle 13 that directs the formation of precise target gene domains such as *hunchback* and *knirps*. (**B**) Proposed gene regulatory network showing cross-regulation of target genes (*9*). The four genes are also under control of Bicoid and other players. *t*, time.

In this Review, we hope to introduce scientists familiar with computational methods (geeks) to a selected set of interesting developmental problems (Fig. 1) and to illustrate to developmental biologists (nerds) a selected set of powerful tools. We focus on recent studies investigating four developmental patterning strategies: (i) gradients of signaling molecules released from localized source cells that guide global patterns across target cell populations (Fig. 1A). This external control contrasts with self-organizing strategies within the cell population that use local interactions, such as (ii) activator-inhibitor mechanisms (Fig. 1B) and (iii) the synchronization of cellular oscillations (Fig. 1C). (iv) Mechanical deformations can also change the pattern of a cellular population (Fig. 1D). Although models are often useful in explaining and predicting developmental phenomena, the eventual fate of a given model is to be proven wrong and then modified or replaced, as illustrated in the companion article on cell polarity by Mogilner and colleagues on page 175 of this special issue. Perhaps the greatest impact of computational approaches in developmental biology right now is to force hypotheses to be precisely stated and to stimulate corresponding new quantitative experiments to test them.

Patterning with Signaling Gradients

Morphogens are diffusible signaling molecules that can activate target genes in a concentration-dependent manner. During development, morphogen gradients are established across tissues, diffusing away from localized sources (Fig. 1A). It has been proposed that cells read morphogen levels to determine their position within the tissue and differentiate accordingly (*6*), and there is good evidence that morphogen gradients can direct cell differentiation in target cells. How these gradients are formed, and whether they are sufficient to control differentiation in very precise domains, are open questions that have benefited from computational approaches.

An important model system for studying these questions is the early embryo of the fruit fly *Drosophila*, in part because its geometry and symmetry simplify description and quantitation (Fig. 2A). One of the maternally deposited cues that breaks the symmetry along the embryo's long axis is *bicoid* mRNA, which is present only in the anterior pole. Bicoid protein is translated and transported (*7*), creating within an hour an exponentially decreasing concentration gradient over several hundred micrometers along the embryo's axis. This gradient directs the formation of precise domains of four target genes—among them *hunchback*—that establish the first segments of the future fly body (Fig. 2A). Given the stochastic nature of gene expression, discussed in the companion article by Munsky and colleagues on page 183 of this special issue, morphogen concentration is expected to fluctuate, both over developmental time and from one individual to another. The stunning precision in the position of the boundaries of the segmented out-

Downloaded from http://science.sciencemag.org/ on December 21, 2016

put pattern that is found despite these fluctuations puzzles both nerds and geeks. The field has wrestled with the issue of whether this precision can be achieved through the Bicoid gradient alone, or whether other mechanisms are required.

Contributing to this debate, recent papers by Manu *et al*. (*8*, *9*) formulated the interactions between four target genes downstream of the maternal gradients in the early embryo using a gene regulatory network (GRN) model, in which each variable represents the quantity of a molecular species (Fig. 2B). One of the limitations of GRN models is that great experimental effort is often required to estimate relevant values of the model's many parameters in the embryo. Parameters for this *Drosophila* segmentation model were obtained computationally by finding those combinations that best reproduced a time series of quantitative spatial gene expression data from the embryo. The model hinted that cross-regulatory interactions between target genes in the GRN reduce the variability in the position of their expression domains.

One problem in understanding a model is that as the parameters vary, the general dynamic behavior of the system can change dramatically. These changes are called bifurcations, and using powerful tools from dynamical systems theory (*10*), Manu *et al*. (*9*) performed a bifurcation analysis of the model to identify the fundamental behaviors that the system can display over a given set of realistic parameter values. The model predicts that cells in the anterior of the embryo select a stable state of the dynamics, and the concentrations of targets change as Bicoid levels drop. In the posterior of the embryo, the system never reaches a stable state because gastrulation happens first. Describing the simple behaviors of a complex regulatory network in this compact way is appealing because it makes similarities to other regulatory systems clearer and also makes falsifiable predictions about distinctive behaviors that can be experimentally tested.

Fluctuations in gene product levels generate molecular noise that limits the precision of signaling gradients and also degrades the targets' outputs. This problem can be formulated precisely by using the tools and concepts from information theory—originally used in engineering—which quantifies the flow of information through communication channels. A key concept is the mutual information between two variables, such as, for example, Bicoid and Hunchback levels. An elegant computation by Tkačik and Walczak used existing precise measurements of morphogen levels (*11*) to estimate the mutual information between Bicoid and Hunchback (*12*). On the basis of their result, they argued that if similar results hold for the other target genes under Bicoid control, the combined information conveyed by the four genes would be enough so that each of the roughly 100 rows of nuclei could unambiguously determine its position along the *Drosophila* embryo. To test this hypothesis, combined high-quality spatial expression data for the other target genes in the system will be necessary. Thus, information theory is emerging as a potentially powerful tool to quantify information transmission in developmental GRNs. As yet, it is unclear whether the *bicoid* gradient is sufficiently precise to instruct the precise boundaries of its target gene domains, or whether other mechanisms are necessary, but computational biology has a central role in this discussion.

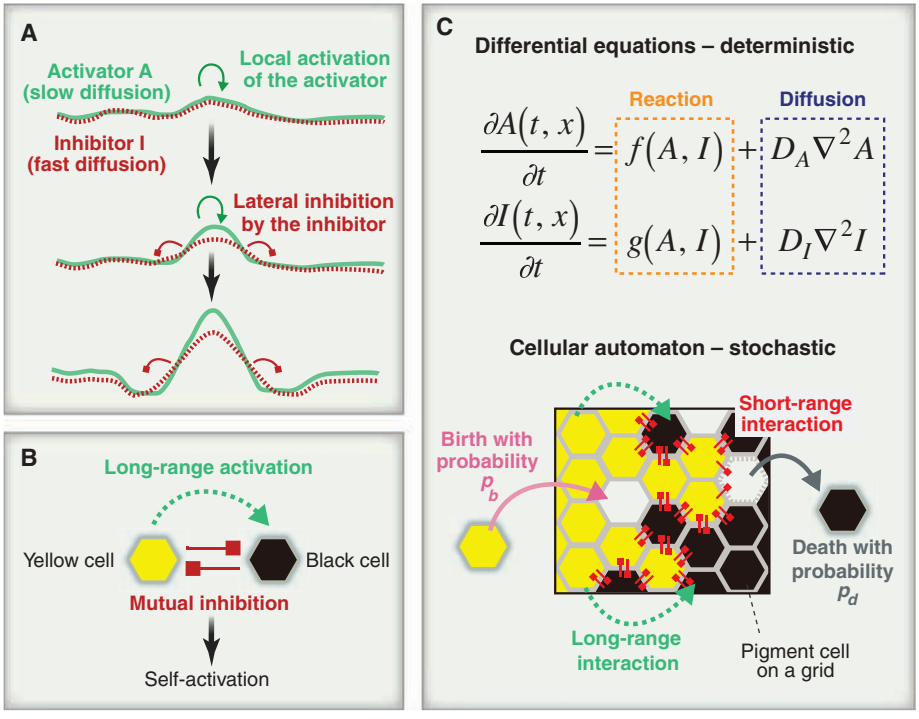


Fig. 3. Patterning with activator-inhibitor systems. (**A**) Local activation and lateral inhibition generates spatially heterogeneous patterns. (**B**) Interactions between black and yellow pigment cells produce Turing patterns in zebrafish skin. Mutual inhibition between them functions as self-activation for the yellow cells. Each yellow cell activates distant black cells. Therefore, inhibition of the yellow cell by the black cell works as a lateral inhibition. (**C**) Different modeling approaches to spontaneous pattern formation.

Patterning with Activator-Inhibitor Systems

Cells in a morphogen gradient use the local level of an externally provided signal to produce patterns (Fig. 1A). However, patterns such as spots and stripes can arise spontaneously from entirely local interactions. In 1952, Alan Turing proposed a reaction-diffusion (RD) mechanism to explain spontaneous pattern formation without signaling gradients (*13*). Specifically, he considered two diffusing chemical components, an activator and an inhibitor (Figs. 1B and 3A). By self-activation, the activator can locally increase its concentration (Fig. 3A). The activator in that region produces the inhibitor, which suppresses the activator in surrounding space because of faster diffusion. As a result, local peaks of activator self-organize from the almost homogeneous starting state, leading to the spontaneous formation of spatial patterns, such as stripes and spots in a two-dimensional (2D) space (so-called Turing patterns) (Fig. 1B).

Subsequently, RD systems have been considered to play important roles in spontaneous pattern formation (*14*, *15*). Although spatial structures very similar to simulated Turing patterns have been observed in development, until recently there was scant evidence showing that the Turing mechanism causes these structures. Indeed, conceptually elegant RD models of the *Drosophila* segmentation process introduced above proved to be entirely wrong (*16*), and this failure may even have left some developmental biologists wary of further theoretical efforts. However, identification of interaction rules and key molecular components in several putative RD systems (*17*, *18*) now suggests the potential of a long-awaited experimental verification of these ideas.

Skin pattern formation in fish has long been a candidate for patterning by use of the Turing mechanism (*19*). To identify key interaction rules in the system, Nakamasu *et al*. studied stripe formation in zebrafish skin (*20*). These black and yellow stripes are self-organized over 3 weeks by local interactions between black and yellow pigment cells, which fulfill the condition for Turing patterns (Fig. 3B). To confirm that the experimentally observed interactions between pigment cells can generate stripes, the authors first used deterministic partial differential equations to model cellular dynamics. However, because the width of each stripe in zebrafish is only ~10 cells, Nakamasu *et al*. pointed out that stochastic effects caused by smaller cell numbers might prevent stable stripe formation. In that situation, it would

Downloaded from http://science.sciencemag.org/ on December 21, 2016

be a better formulation to explicitly describe stochastic behaviors of each single pigment cell, such as birth, movement, and cell death. The authors developed a cellular automaton-based model (Fig. 3C) that includes the observed pigment cell interactions to study the robustness of stripe patterns against stochastic effects. Although such detailed models usually include several parameters not measured experimentally, simulations of the cell-based model produced patterns similar to those obtained by the deterministic model and observed on the zebrafish skin. Combining investigations of the molecular and cellular basis of the cellular-level interaction rules (*21*) with further theoretical studies should reveal whether this is indeed a Turing system.

Gradient patterning strategies can also be formulated as RD systems because gradients can arise from diffusion of morphogens, and the pattern emerges due to reactions that involve these morphogens. However, the different length-scales involved in activator-inhibitor systems give rise to qualitatively different patterns, which are local in nature. This is an example of how very different developmental patterning strategies can be described by using similar model formulations.

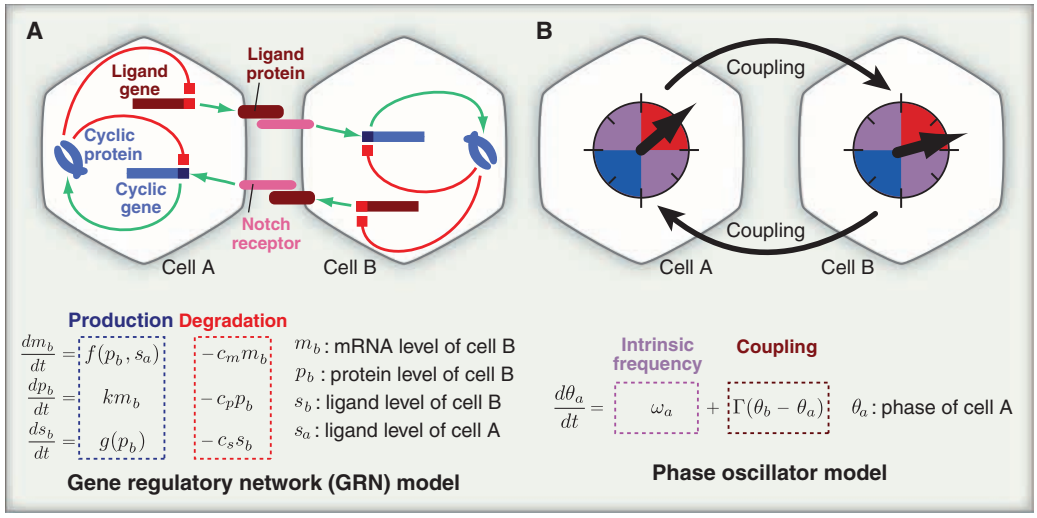


Fig. 4. Patterning with genetic oscillations. (**A**) Cyclic gene expression oscillates in individual cells because of a negative feedback loop, and oscillations are coupled to neighbor cells through the Notch pathway. (**B**) The mutual effects of cellular oscillators can be described by models of coupled phase oscillators.

Patterning with Genetic Oscillations

The growing body axis of all vertebrate embryos is rhythmically and sequentially subdivided into segments. For example, in the zebrafish embryo the multicellular segments are ~50 μm long and form with a periodicity of 30 min. Inspired by such clock-like regularity, Cooke and Zeeman proposed the Clock and Wavefront model in 1976 (*22*). In this model, a biological clock ticks at the posterior of the elongating embryo, and the distance advanced by a wavefront along the embryonic axis during a cycle of the clock sets the length of a forming segment. More than 20 years later, the model was revived with the discovery of genetic oscillations in the chick embryo (*23*). This segmentation clock appears to be a tissue-level rhythmic pattern generator (*24*), in which a population of progenitor cells behave as coupled oscillators, self-organizing a collective rhythm through mutual synchronization (Fig. 1C).

A clue to the existence of such a synchronized cell population came from zebrafish mutants that disrupt Delta-Notch intercellular signaling, in which coherent oscillations and segmental patterning are gradually lost (*25*). The current hypothesis is that in the wild-type embryo, Delta ligands under the control of a single-cell oscillator activate Notch receptors in the membrane of neighboring cells, and these receptors coordinate oscillating gene expression in the receiving cell (Fig. 4A). Without Delta-Notch signaling, the single cells' oscillations gradually lose synchrony. The plausibility of this synchronization hypothesis has been studied by using GRN models showing that the Delta-Notch mechanism described above could keep neighboring cells oscillating in synchrony (*26*, *27*).

Given the previously mentioned difficulty of determining GRN parameters from embryos (*28*), an alternative and complementary model formulation is to use an effective theory with variables that represent processes for which there is a particular interest or a possibility of experimental comparison. For the segmentation clock, this approach has been applied to investigate the synchronization hypothesis by using theories based on coupled phase oscillators (Fig. 4B). In a phase oscillator model, the variables corresponding to oscillating molecular species are substituted by a single variable: the phase of the oscillation cycle, which advances in time with a given intrinsic frequency. The effect of Delta-Notch signaling is captured by a coupling function that speeds up or slows down a cellular oscillator depending on the phase of neighboring cells. Phase oscillator models do not offer direct insight about dynamics of individual molecular species, but their simplicity allows powerful insights about system-level dynamics from paper-and-pencil analysis. Furthermore, they allow a direct fit to experimental data relying on a few coarse-grained parameters such as the period of the oscillations (*29*).

Using a phase oscillator model, the synchronization problem of the segmentation clock was formulated as a competition between noise and the intercellular coupling that keeps cells in synchrony (*30*). Together with quantitative experimental disruptions of Notch signaling in zebrafish, the model allowed estimation of the noise level and coupling strength relevant for the tissue-level synchrony of the clock. Coupling involves the new synthesis of Delta ligand every cycle (Fig. 4A), and to represent the anticipated duration of the ligand-receptor mechanism, Morelli *et al.* (*29*) included explicit time delays in the coupling function of a phase oscillator model. This delayed coupling theory made the prediction that changing the coupling strength could change the clock period and motivated the study of the dynamics of Notch mutants. Quantitative time-lapse measurements of segmentation period and analysis of clock gene-expression patterns in mutants matched the theoretical predictions and so identified the first candidates for segmentation clock period mutants (*31*).

Although these studies have revealed some surprising insights into the segmentation clock's dynamics, most quantitative data used to test models have come from static images (*28*, *31*), and the desynchronization of the clock has not been directly observed. The advent of new techniques to observe cyclic gene expression in vivo (*32*) will allow key assumptions of the existing models to be directly tested.

Patterning with Mechanical Deformations

We complete our roster of patterning mechanisms with a recently discovered case driven by tissue deformations. An apparently simple behavior for an epithelial sheet is to elongate along one axis while shrinking along the orthogonal axis. During *Drosophila* development, the wing blade epithelium stretches into the familiar elongate wing shape, and each of the hairs protruding from the wing cells points distally—an example of planar cell polarity (PCP) patterning (Fig. 5A). Although proximo-distal gradients of PCP pathway components have been observed, they are not sufficient to produce the final wing hair polarity (*33*). Examination of cell shapes and trajectories from time-lapse movies shows that sharp contraction of the neighboring hinge region exerts anisotropic tension on the wing blade (*34*). Over a period of 15 hours, the blade deforms with a shear gradient arising from the cellular flow in the tissue.

Downloaded from http://science.sciencemag.org/ on December 21, 2016

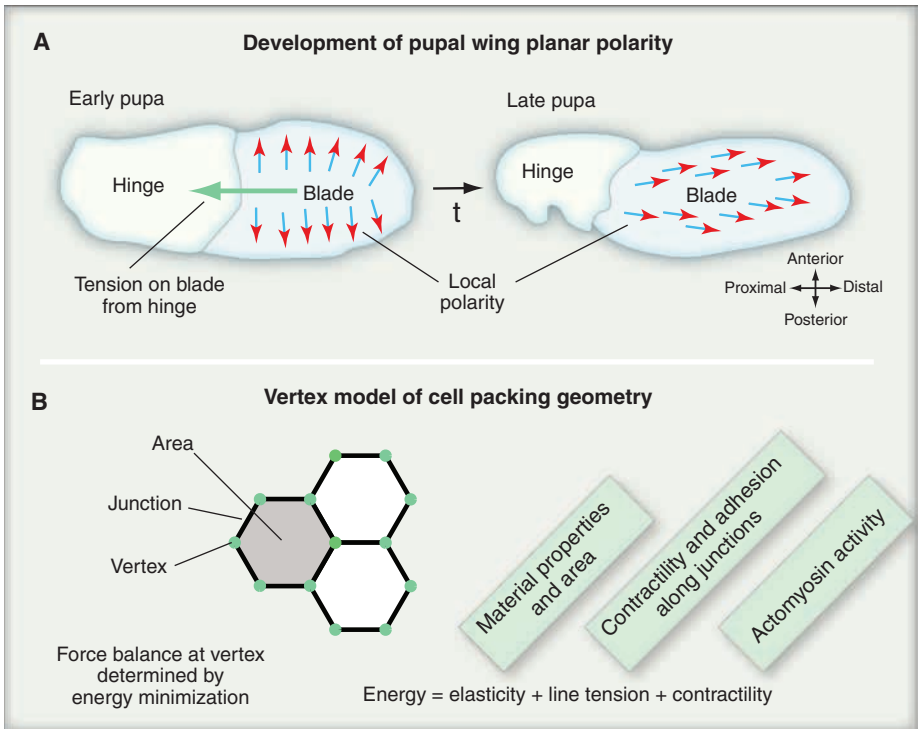


Fig. 5. Patterning by mechanical deformation. (**A**) Overview of *Drosophila* wing development during pupariation, when the wing blade elongates and proximo-distal planar polarity is established. (**B**) Schematic of the vertex model used to calculate stable cell-packing geometries.

Aigouy *et al*. explored the role of tissue shear in aligning the axis of cellular polarity with the proximo-distal axis of the wing blade by formulating a 2D vertex model of epithelial cell shape (Fig. 5B) (*35*), incorporating an effective description of the local recruitment of complementary PCP molecules to apposing cell boundaries (*34*). This new model predicts that polarity is reoriented by local rotation and cell flow–induced shear. Simulations show that shear associated with oriented cell division, proximo-distal cell elongation, and cell rearrangement also contribute to the alignment of cell polarity with the long axis of the wing. Future work can investigate how the 3D baso-lateral surfaces of the epithelial cells in the wing affect this description, and how the PCP protein complexes involved dynamically reorganize during cellular rearrangement. Thus, remarkably the final planar cell polarity of the completed wing may be a direct consequence of the externally applied stresses responsible for its extension, via simple physical rules such as those that determine molecular polarity in liquid crystals (*36*).

In this Review, we have mainly discussed chemical aspects of pattern formation as separate from downstream mechanics of morphogenesis (*37*, *38*). Turing already wondered whether a closer linkage might be at work (*13*), and it seems timely to reconsider development as having integrated mechanochemical aspects (*39*). For example, motivated by recent findings on cell cortex dynamics in the nematode *Caenorhabditis* (*40*), Bois *et al*. studied pattern formation in an active fluid in which mechanical contraction causes the flow of reactive chemical species (*41*). This theoretical analysis showed that an active fluid extends the parameter space in which classical Turing systems generate spatial patterns. To what extent continuous feedback between chemical and mechanical processes also underlies tissue-level phenomena in development is not yet clear, but it may be widespread.

Outlook

With the wide range of approaches in use, how should the developmental biologist select the appropriate modeling and computational methods? And where should the computational scientist dig for interesting problems in the vast field of developmental biology? Previous reviews have given multiple examples and advice (*1*–*5*). Here, we argue that the first step is key: The level of description and model type should be matched to the best available data. The data should be quantitative, accurate, and precise, and the model should make falsifiable predictions. Although some researchers are fluent in both domains, most often a successful computational approach to developmental biology will involve a long-term dialogue between experts across disciplinary boundaries. As advances in imaging and molecular methods increase experimental resolution and complexity, corresponding theoretical and computational developments will be required to assemble the puzzle. This co-dependence should generate a wealth of new opportunities for geeks and nerds alike.

References and Notes

1. A. C. Oates, N. Gorfinkiel, M. González-Gaitán, C. P. Heisenberg, *Nat. Rev. Genet.* **10**, 517 (2009).
2. J. Lewis, *Science* **322**, 399 (2008).
3. G. T. Reeves, C. B. Muratov, T. Schüpbach, S. Y. Shvartsman, *Dev. Cell* **11**, 289 (2006).
4. C. J. Tomlin, J. D. Axelrod, *Nat. Rev. Genet.* **8**, 331 (2007).
5. S. Roth, *Dev. Genes Evol.* **221**, 255 (2011).
6. L. Wolpert, *J. Theor. Biol.* **25**, 1 (1969).
7. S. C. Little, G. Tkačik, T. B. Kneeland, E. F. Wieschaus, T. Gregor, *PLoS Biol.* **9**, e1000596 (2011).
8. Manu *et al*., *PLoS Biol.* **7**, e1000049 (2009).
9. Manu *et al*., *PLOS Comput. Biol.* **5**, e1000303 (2009).
10. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview, Boulder, CO, 1994).
11. T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, D. W. Tank, *Cell* **130**, 141 (2007).
12. G. Tkačik, A. M. Walczak, *J. Phys. Condens. Matter* **23**, 153102 (2011).
13. A. M. Turing, *Philos. Trans. R. Soc. London Ser. B* **237**, 37 (1952).
14. H. Meinhardt, *Models of Biological Pattern Formation* (Academic Press, London, 1982).
15. J. D. Murray, *Mathematical Biology* (Springer Verlag, Berlin, 2003).
16. M. Akam, *Nature* **341**, 282 (1989).
17. A. D. Economou *et al*., *Nat. Genet.* **44**, 348 (2012).
18. M. V. Plikus *et al*., *Science* **332**, 586 (2011).
19. A. Kondo, R. Asai, *Nature* **376**, 765 (1995).
20. A. Nakamasu, G. Takahashi, A. Kanbe, S. Kondo, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8429 (2009).
21. M. Inaba, H. Yamanaka, S. Kondo, *Science* **335**, 677 (2012).
22. J. Cooke, E. C. Zeeman, *J. Theor. Biol.* **58**, 455 (1976).
23. I. Palmeirim, D. Henrique, D. Ish-Horowicz, O. Pourquié, *Cell* **91**, 639 (1997).
24. A. C. Oates, L. G. Morelli, S. Ares, *Development* **139**, 625 (2012).
25. Y. J. Jiang *et al*., *Nature* **408**, 475 (2000).
26. J. Lewis, *Curr. Biol.* **13**, 1398 (2003).
27. K. Uriu, Y. Morishita, Y. Iwasa, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4979 (2010).
28. F. Giudicelli, E. M. Ozbudak, G. J. Wright, J. Lewis, *PLoS Biol.* **5**, e150 (2007).
29. L. G. Morelli *et al*., *HFSP J.* **3**, 55 (2009).
30. I. H. Riedel-Kruse, C. Müller, A. C. Oates, *Science* **317**, 1911 (2007).
31. L. Herrgen *et al*., *Curr. Biol.* **20**, 1244 (2010).
32. Y. Niwa *et al*., *Genes Dev.* **25**, 1115 (2011).
33. D. Ma *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18800 (2008).
34. B. Aigouy *et al*., *Cell* **142**, 773 (2010).
35. R. Farhadifar, J. C. Röper, B. Aigouy, S. Eaton, F. Jülicher, *Curr. Biol.* **17**, 2095 (2007).
36. J. F. Joanny, F. Julicher, K. Kruse, J. Prost, *New J. Phys.* **9**, 422 (2007).
37. N. Gorfinkiel, S. Schamberg, G. B. Blanchard, *Genesis* **49**, 522 (2011).
38. S. M. Trier, L. A. Davidson, *Curr. Opin. Genet. Dev.* **21**, 664 (2011).
39. J. Howard, S. W. Grill, J. S. Bois, *Nat. Rev. Mol. Cell Biol.* **12**, 392 (2011).
40. M. Mayer, M. Depken, J. S. Bois, F. Jülicher, S. W. Grill, *Nature* **467**, 617 (2010).
41. J. S. Bois, F. Jülicher, S. W. Grill, *Phys. Rev. Lett.* **106**, 028103 (2011).

Acknowledgments: We thank F. Jülicher, C.-P. Heisenberg, S. Grill, P. R. ten Wolde, T. Bollenbach, P. Formosa-Jordan, and members of the Oates and Jülicher groups for discussion and critical comments. This work was supported by the Max Planck Society and the European Research Council (ERC) under the European Communities Seventh Framework Programme (FP7/2007-2013)/ERC grant 207634. K.U. is supported by the Japan Society for the Promotion of Science for Young Scientists. S.A. acknowledges funding from CSIC through the "Junta para la Ampliación de Estudios" program (JAEDOC014, 2010 call) co-funded by the European Social Fund, and from Ministerio de Ciencia e Innovación (Spain) through MOSAICO.

10.1126/science.1215478

Downloaded from http://science.sciencemag.org/ on December 21, 2016

EXTENDED PDF FORMAT
SPONSORED BY

www.rndsystems.com

Computational Approaches to Developmental Patterning

Luis G. Morelli, Koichiro Uriu, Saúl Ares and Andrew C. Oates

(April 12, 2012)
Science **336** (6078), 187-191. [doi: 10.1126/science.1215478]

Editor's Summary

This copy is for your personal, non-commercial use only.

Article Tools Visit the online version of this article to access the personalization and article tools:
http://science.sciencemag.org/content/336/6078/187

Permissions Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Downloaded from http://science.sciencemag.org/ on December 21, 2016

frontiers in Bioengineering and Biotechnology

PERSPECTIVE
published: 26 June 2020
doi: 10.3389/fbioe.2020.00705

Toward Engineering Biosystems With Emergent Collective Functions

Thomas E. Gorochowski[1], Sabine Hauert[2†], Jan-Ulrich Kreft[3†], Lucia Marucci[2†], Namid R. Stillman[2†], T.-Y. Dora Tang[4,5†], Lucia Bandiera[6], Vittorio Bartoli[2], Daniel O. R. Dixon[7], Alex J. H. Fedorec[8], Harold Fellermann[9], Alexander G. Fletcher[10], Tim Foster[3], Luca Giuggioli[2], Antoni Matyjaszkiewicz[11], Scott McCormick[2], Sandra Montes Olivas[2], Jonathan Naylor[9], Ana Rubio Denniss[2] and Daniel Ward[1]*

[1] School of Biological Sciences, University of Bristol, Bristol, United Kingdom, [2] Department of Engineering Mathematics, University of Bristol, Bristol, United Kingdom, [3] School of Biosciences and Institute of Microbiology and Infection and Centre for Computational Biology, University of Birmingham, Birmingham, United Kingdom, [4] Max Plank Institute of Molecular Cell Biology and Genetics, Dresden, Germany, [5] Physics of Life, Cluster of Excellence, Technische Universität Dresden, Dresden, Germany, [6] School of Engineering, University of Edinburgh, Edinburgh, United Kingdom, [7] School of Biochemistry, University of Bristol, Bristol, United Kingdom, [8] Division of Biosciences, University College London, London, United Kingdom, [9] School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom, [10] Bateson Centre and School of Mathematics and Statistics, University of Sheffield, Sheffield, United Kingdom, [11] The European Molecular Biology Laboratory, Barcelona, Spain

OPEN ACCESS

Edited by:
Pablo Ivan Nikel, Novo Nordisk Foundation Center for Biosustainability (DTU Biosustain), Denmark

Reviewed by:
Dae-Hee Lee, Korea Research Institute of Bioscience and Biotechnology (KRIBB), South Korea
Irene Otero Muras, Consejo Superior de Investigaciones Científicas (CSIC), Spain

***Correspondence:**
Thomas E. Gorochowski
thomas.gorochowski@bristol.ac.uk

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *02 May 2020*
Accepted: *05 June 2020*
Published: *26 June 2020*

Citation:
Gorochowski TE, Hauert S, Kreft J-U, Marucci L, Stillman NR, Tang T-YD, Bandiera L, Bartoli V, Dixon DOR, Fedorec AJH, Fellermann H, Fletcher AG, Foster T, Giuggioli L, Matyjaszkiewicz A, McCormick S, Montes Olivas S, Naylor J, Rubio Denniss A and Ward D (2020) Toward Engineering Biosystems With Emergent Collective Functions. Front. Bioeng. Biotechnol. 8:705. doi: 10.3389/fbioe.2020.00705

Many complex behaviors in biological systems emerge from large populations of interacting molecules or cells, generating functions that go beyond the capabilities of the individual parts. Such collective phenomena are of great interest to bioengineers due to their robustness and scalability. However, engineering emergent collective functions is difficult because they arise as a consequence of complex multi-level feedback, which often spans many length-scales. Here, we present a perspective on how some of these challenges could be overcome by using multi-agent modeling as a design framework within synthetic biology. Using case studies covering the construction of synthetic ecologies to biological computation and synthetic cellularity, we show how multi-agent modeling can capture the core features of complex multi-scale systems and provide novel insights into the underlying mechanisms which guide emergent functionalities across scales. The ability to unravel design rules underpinning these behaviors offers a means to take synthetic biology beyond single molecules or cells and toward the creation of systems with functions that can only emerge from collectives at multiple scales.

Keywords: synthetic biology, multi-agent modeling, systems biology, emergence, multi-scale, bioengineering, consortia, collectives

INTRODUCTION

Many living organisms have evolved traits to exploit the capabilities that emerge from large interacting populations of molecules or cells, which go beyond those of the individual elements. From bacteria forming biofilms to fight antibiotic treatments to synchronizing their behaviors through quorum sensing during disease, emergent collective behaviors are pervasive in biology. Likewise, the engineering of emergent collective behaviors could offer an intriguing path to artificial biosystems with improved reliability, robustness and scalability. However, current approaches to biological design are ill-equipped for this task as they tend to focus on a single level of

organization and ignore potential feedbacks between different aspects/levels of a system. A common example is the design of transcriptional gene regulatory networks where it is assumed that the function of the entire system can be understood solely by the steady state input-output transcriptional response of genetic devices (Nielsen et al., 2016). While this simplification is useful and powerful in some cases, if the genes regulated link to metabolic processes there is a chance that feedback via metabolism could break circuit function. Focusing purely on transcriptional networks makes it impossible to capture such behaviors.

In physics, great strides have been made through techniques from statistical mechanics to understand emergent phenomena. These include the Ising model used to capture magnetic phase transitions (Taroni, 2015) and the application of renormalization to understand how physical and biological constraints might underpin scaling laws that guide evolution (West et al., 2002; Kempes et al., 2019). There has also been growing interest over the past few decades in the field of complexity theory (Nicolis and Prigogine, 1989) and whether laws might exist that govern self-organization and emergence across diverse types of complex system composed of many interacting parts (Prigogine and Nicolis, 1985; Ashby, 1991; Goldstein, 1999; West et al., 2002).

An approach to capture and explore the emergent features of complex systems is multi-agent modeling (also termed agent-based or individual-based modeling) (Hellweger et al., 2016). This considers key components of a system as explicit entities/agents and allows for large and diverse interacting populations of these (**Figure 1A**). Specifically, a multi-agent model consists of autonomous agents that represent the lowest level components of the system. Common types of agent in biological systems include molecules, cells and whole multicellular organisms. Each agent is assigned a specific set of rules governing how it interacts with other agents and the local environment. The way these rules are modeled is flexible with the option to use basic finite state-machines, Boolean logic governing stimuli-response relationships, or more complex representations like differential equation models (e.g., capturing the biochemical reaction networks within a cell). Populations of these agents are then placed in a simulated environment that encompasses physical processes of relevance to the system. In biology, this might include the diffusion of chemicals, the flow of fluids, and the mechanical forces that cells can exert on one another. Again, the way that these environmental processes are modeled can vary, resulting in a final model that could potentially combine stochastic, deterministic, dynamic, discrete and continuous representations for different aspects of a system. The integration of such diverse modeling approaches allows for the most appropriate form of representation to be used for each aspect and helps simplify the specification of the multi-scale system, but often comes at the cost of reduced analytical tractability. Even so, multi-scale modeling has been shown capable of discovering some of the core ingredients needed for collective behaviors to emerge (Hellweger et al., 2016; Gorochowski and Richardson, 2017), but its use to date in synthetic biology has been limited (Gorochowski, 2016).

Here, we aim to highlight some of the key areas of synthetic biology where multi-agent modeling offers a unique way to tackle longstanding problems (**Figure 1B**). While the examples we cover are diverse, they all share a core characteristic: the emergence of behaviors in the systems cannot be explained by looking solely at their basic parts in isolation. This essence makes such systems special yet difficult to engineer via traditional means. We propose to extend bioengineering methods to encompass principles gleaned from multi-agent models and use them to guide the design of synthetic biological systems displaying emergent phenomena. We end by discussing some of the practical challenges when using multi-agent modeling in synthetic biology and future directions for the marriage of these fields.

UNDERSTANDING THE EMERGENCE OF LIFE

When considering emergent phenomena, the quintessential example is the emergence of life. Putting aside the difficulty of defining precisely what life is, the ability of living systems to self-replicate and create order/information out of chaos is an inspiration for many engineers. Bottom-up synthetic biology attempts to build chemical systems that display life-like behaviors using a minimal set of components. The hope is that these simplified systems might help us understand how life emerged from first principles.

One attempt to reach this goal has been via the synthesis of artificial cells (protocells) with life-like properties. This requires the ability to bridge length scales by harnessing molecular self-assembly to create micron-sized compartments (Bayley et al., 2008; Li et al., 2014) and the intricate interactions between molecules and enzymes to form biochemical reaction networks (Hasty et al., 2002). The incorporation of these reaction networks within protocells has also been demonstrated (Adamala et al., 2017; Joesaar et al., 2019) and although chemically simple, such systems display an array of dynamical behaviors including pattern formation (Niederholtmeyer et al., 2015; Zadorin et al., 2017) and replication via controlled growth and division (Chen et al., 2004). By combining these systems with additional chemical modules and parts, this may offer a route to creating other key behaviors of living systems.

Building on these capabilities, functionalities can be scaled further by constructing systems composed of populations of protocells or through interacting natural and artificial cellular communities (Lentini et al., 2014; Adamala et al., 2017; Tang et al., 2018). While such extensions offer a promising platform for probing emergent behaviors using simple self-contained chemical units, it is difficult to know what parameters to engineer into these systems and the level of complexity required to drive a desired collective behavior. This is where multi-agent modeling, in combination with more traditional models of chemical reaction systems, could lead to a quantitative understanding of the key elements needed for the emergence of life-like behaviors. In particular, multi-agent models would allow for the rapid exploration of potential systems using physically

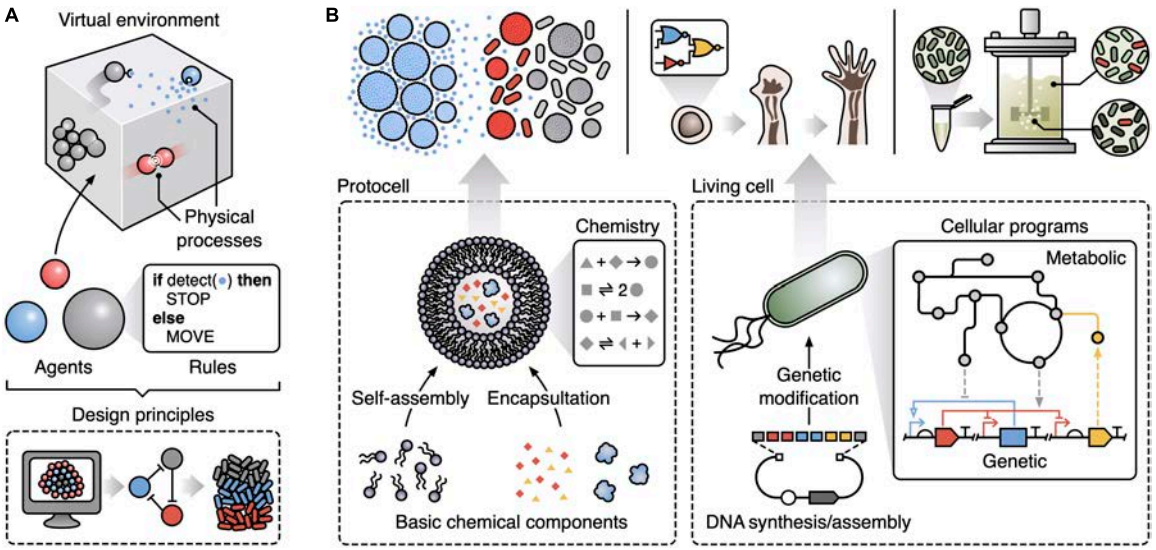


FIGURE 1 | Multi-agent modeling can support the design of emergent collective functions in synthetic biology. **(A)** Key components of a multi-agent model. Populations of autonomous agents following user-prescribed rules are placed in a virtual environment that simulates relevant physical processes (e.g., physical collisions, chemical diffusion, movement, and fluid flows). Simulations of multi-agent models can be used to derive design principles that capture the basic ingredients (e.g., specific patterns of interactions) needed for a particular emergent behavior. **(B)** Potential applications of multi-agent modeling within synthetic biology and the underlying agents (bottom, dashed boxes) used to generate specific emergent collective behaviors: (top left) exploring how to create life-like behaviors from basic chemical components with sender protocells (blue) able to spatially propagate a signal to receiver protocells and bacteria (gray when inactive, red when active) using a small diffusive chemical (small blue dots); (top middle) understanding the developmental programs used during morphogenesis as a step toward the creation of synthetic multi-cellular life; (top right) improving scale-up of microbial fermentations by accounting for heterogeneity across a bioreactor and designing engineered microbes able to robustly function under these conditions.

realistic parameters until the right combination of parts was found that resulted in a desired emergent functionality.

Historically, mathematical models developed using differential equations have proved effective for understanding the dynamics of minimal chemical systems (Rovinskii and Zhabotinskii, 1984) and are widely and successfully used for modeling all types of biological system (Ellner and Guckenheimer, 2011; Raue et al., 2013). Furthermore, the application of bifurcation analysis to these dynamical models enables the rigorous characterization of emergent phenomena such as bi-stability, symmetry breaking, non-linear oscillations, chaos, and pattern formation (Kuznetsov, 2004). However, while it is possible to use partial differential equations (PDEs) to capture spatial aspects of a system, the high levels of heterogeneity in the complex environments of many biological system (e.g., cellular tissues) and the ability of both agents and the rules to change over time, can make practical use of PDEs a challenge (Hellweger et al., 2016; Perez-Carrasco et al., 2016; Glen et al., 2019).

In comparison, multi-agent modeling is able to explicitly capture such variation and consider simplified rules to express internal chemical reactions altering specific characteristics of each component. Due to the chemical simplicity and programmability of minimal protocells, this abstraction is a good fit, allowing iterative refinement of model and experimental system. For example, due to the limited number of possible chemical reactions present in a minimal system, comprehensive direct measurements can be made to create highly predictive rules for how a protocell's chemical state will change over time. These can then drive simulations of accurate protocell behaviors in a multi-agent model to explore the specific combination of reactions required for the emergence of higher population-level functionalities. This two-way cycle of development would be difficult, if not impossible, when using natural cells where complex evolutionary baggage masks those features essential for emergence.

DISTRIBUTED COMPUTATION DURING DEVELOPMENT

Living cells continually monitor their environment and adapt their physiology in order to survive. This requires the processing of information gathered from sensors to make suitable changes to gene expression. Synthetic biology enables us to reprogram cells by writing our own genetic programs to exploit the cells' computational capabilities in new ways (Greco et al., 2019; Grozinger et al., 2019). So far, the majority of research in biological computation has revolved around the concept of genetic circuits and attempted to repurpose tools and methodologies from electronic circuit design (Nielsen et al., 2016; Gorochowski et al., 2017) and automatic verification (Dunn et al., 2014). While this approach has enabled the automated design of cellular programs able to perform basic logic, much of the information processing in native biological systems is distributed,

relying on collective decision making (e.g., quorum sensing) and interactions between large numbers of cells.

This feature is most evident in developmental biology where robust genetic programs must ensure that a complex multi-cellular organism emerges from a single cell. Cell growth, differentiation, migration and self-organization are coordinated by a developmental program with dynamics at both the intra- and inter-cellular levels. These enable the generation of precise deterministic patterns from stochastic underlying processes (Glen et al., 2019). In contrast to simple logic circuits, the complexity of the molecular interactions and mechanical forces underpinning these processes motivate the use of multi-agent modeling to better understand how developmental programs work in morphogenetic systems. In particular, multi-agent models are able to capture the role of cellular heterogeneity, proliferation and morphology, mechanical and environmental cues, movement of cells as well as the integration of multiple processes at diverse scales and the feedback between these (Montes-Olivas et al., 2019). Such models have helped deepen our understanding of early mammalian embryogenesis (Godwin et al., 2017), as well as the formation of vascular networks (Perfahl et al., 2017) and other complex structures and organs, including the skin, lung (Stopka et al., 2019), kidney (Lambert et al., 2018), and brain (Caffrey et al., 2014).

Although such work has provided insights into the computational architecture of native developmental programs, it has been difficult to apply this information to the creation of *de novo* morphogenetic systems because of a limited toolkit of parts available to build such systems. Synthetic biology may help solve this issue by facilitating the engineering of simplified multi-cellular systems (Velazquez et al., 2018) that implement developmental programs encompassing distributed feedback regulation (Ausländer and Fussenegger, 2016) and cell-to-cell communication (Bacchus et al., 2012), to better understand how these factors can be used to contribute to emergent self-organization (Morsut et al., 2016).

COLLECTIVE PHENOMENA DRIVING DISEASE

Many of the challenges treating diseases result from the malfunction of emergent multi-cellular properties, be it carcinogenesis (Deisboeck and Couzin, 2009; Ward et al., 2020), viral infection (Jacob et al., 2004), bacterial biofilm formation (Wu et al., 2020) or microbiome imbalances (Shreiner et al., 2015; Kumar et al., 2019). Multi-agent modeling of these conditions has helped demystify how the collective behavior of large numbers of diverse cells and their interactions with each other and their environment can lead to negative clinical outcomes.

Cancer is a complex multi-scale disease that includes environmental factors, genetic mutations and clonal selection, and complex interactions with the immune and vascular system. As a result, computational models of cancer need to account for many of these factors considering the heterogeneity and interactions of single cells, yet contain sufficient numbers of them to predict emergent phenomena at a tumor scale (Metzcar et al., 2019). Using this approach, multi-agent models have been used to help understand metastasis (Waclaw et al., 2015) and show that cancer cells with stem cell-like properties can be a key determinant in cancer progression with fatal consequences (Scott et al., 2016, 2019).

Beyond understanding the emergence of some diseases, multi-agent models can also identify novel ways of fixing their dynamics by considering how to disrupt cellular behaviors, and their interactions in space and time (Waclaw et al., 2015; Gallaher et al., 2018). Treatments themselves can even be designed to have collective emergent properties. For example, bacteria have already been engineered to use quorum sensing to trigger their delivery of drugs (Din et al., 2016) or they can be controlled using magnetic fields to penetrate cancerous tissue (Schuerle et al., 2019). Other collective behaviors used in cancer nanomedicine include self-assembly of nanoparticles to anchor imaging agents in tumors, disassembly of nanoparticles to increase tissue penetration, nanoparticles that compute the state of a tumor, nanoparticle-based remodeling of tumor environments to improve secondary nanoparticle transport, or nanoparticle signaling of tumor location to amplify the accumulation of nanoparticles in tumors (Hauert et al., 2013; Hauert and Bhatia, 2014).

The emergent properties inherent in many diseases, and the potential to harness such behaviors for new treatments, highlight the need for multi-scale modeling tools. Moreover, with the rapidly expanding field of "systems medicine," integrated modeling pipelines able to predict multi-scale disease dynamics and assess novel synthetic biology treatments via large-scale simulation and machine learning are positioned to revolutionize many areas of medicine (Stillman et al., 2020).

CHALLENGES IN SCALING-UP BIOTECHNOLOGY

The ability for synthetic biology to reprogram cellular metabolisms offers an opportunity to convert cheap substrates (or even waste) into valuable chemicals and materials via microbial fermentation (Nielsen and Keasling, 2016). To make this economically viable, large bioreactors are often used. However, while our use of fermentation stems back millennia (McGovern et al., 2004), we still struggle to reliably scale-up many processes from shake flasks in the lab to industrial-sized bioreactors (Lee and Kim, 2015).

A major reason for this problem is the increasing difficulty and power consumption of mixing or aerating reactors as their volume increases, causing pockets to form where nutrient concentration, temperature, oxygen, pH and other factors differ (Alvarez et al., 2005). As a microbe travels through the bioreactor, it becomes exposed to a wide variety of environments, each causing changes in its physiology. Because the path of each cell is unique, a population of cells will display a wide variety of physiological states. This differs from lab-scale experiments where environments are well-mixed and homogeneous, and causes predictions made from these conditions to significantly deviate from those observed during scale-up.

Capturing the combined environmental and cellular variability present in a large bioreactor is difficult using standard differential-equation models. In contrast, multi-agent models are able to explicitly capture and link gene regulation, metabolism, and the cells' local environment (Nieß et al., 2017; Haringa et al., 2018), as well as differences between individual cells and how cells change over time (González-Cabaleiro et al., 2017). In particular, hybrid models in which continuous descriptions of complex physical processes like fluid flows have been coupled with multi-agent models to allow for the efficient simulation of these systems. This approach can accurately predict the emergence of population heterogeneity and overall production rates and help guide bioreactor design to further improve yields (Haringa et al., 2018). Some attempts have also been made to use control engineering principles to design cellular systems able to adapt to fluctuating environments (Hsiao et al., 2018). To date, these attempts have mostly focused on the basic genetic parts and regulatory motifs (e.g., negative feedback) needed to implement control algorithms (Ceroni et al., 2018; Aoki et al., 2019; Pedone et al., 2019; Bartoli et al., 2020). Moving forward, multi-agent models offer a means to make simulations of these systems more realistic by accurately capturing how individual cells and their complex environment change over time.

Another challenge faced during large-scale fermentation is the opportunity for mutants to arise or unwanted microbes to contaminate a process and out-compete their engineered counterparts (Kazamia et al., 2012; Louca and Doebeli, 2016). Multi-agent models of these complex environments and local competition when multiple types of organism are present, could help support the development of evolutionarily stable strategies (ESSs) that prevent the replacement of an engineered population by competitors (Schuster et al., 2010).

ENGINEERING SYNTHETIC ECOLOGIES

At an even larger organizational level, synthetic biologists have begun to explore how to engineer interactions between communities to enable the future construction of synthetic ecologies (Ben Said and Or, 2017). With climate change, pollution and many other factors leading to the degradation of ecological systems, understanding how these systems emerge and function is crucial. Such knowledge would allow for effective restoration strategies (Solé et al., 2015) and potentially offer means to terraform other planets like Mars for future human inhabitation (Conde-Pueyo et al., 2020).

These applications require an understanding of how diverse organisms interact to create stable communities (Widder et al., 2016). This is difficult because the interactions that take place at the level of a population are governed by choices made by single organisms (Kreft et al., 2017). By using multi-agent modeling to rapidly test combinations of cell types, behaviors and interactions, and synthetic biology tools to engineer real-world microbial communities, it might become possible to design and test hypotheses regarding the principles for robust ecosystem design. For example, multi-agent modeling has been used to help understand how signaling and mutual cooperation can stabilize microbial communities (Kerényi et al., 2013). Furthermore, from a synthetic biology perspective many of the tools needed to engineer these systems already exist, e.g., biological parts able to implement cooperation (Shou et al., 2007), signaling (Bacchus et al., 2012), targeted death (Fedorec et al., 2019), and collective decision making (e.g., quorum sensing).

Beyond engineering interactions between organisms, spatial structure can also play a crucial role in the functionalities of microbial communities. Multi-agent modeling has demonstrated the significant impact that spatio-temporal organization can have on soil microbes and the success of auxotrophic interactions (Jiang et al., 2018). Such interactions are particularly important for engineering minimal functional synthetic communities as plant seed treatments and for vertical farming under defined conditions. In this context, whether or not a single cell or division of labor is the evolutionarily stable solution depends on the metabolic flux through the system, with high flux favoring division of labor (Kreft et al., 2020). Extending this modeling approach further to consider the thermodynamics of microbial growth and redox biochemistry could help ensure that resultant systems are ecologically and evolutionarily stable (Zerfaß et al., 2018). Alternatively, external control of the environment could be used to forcibly maintain a desired community structure (Treloar et al., 2020). In all cases, a combination of multi-agent modeling and engineerable biological systems provides a unique means to unravel how these complex systems function.

External feedback control has been proposed as another approach to control of cellular communities. By employing real-time single cell measurements (e.g., by time-lapse microscopy or flow-cytometry) and experimental systems able to send control signals to the cells via optogenetics (Toettcher et al., 2011) or chemical release in microfluidics (Menolascina et al., 2014), a computer can monitor and signal to a population of cells in order to maintain a desired behavior (e.g., the expression rate of a protein). More recently, it has been proposed to implement these control algorithms directly into cells, with the key aim of distributing tasks among different strains (Fiore et al., 2017; McCardell et al., 2017). Multi-agent modeling can be instrumental in the design of robust feedback mechanisms across multicellular populations, as it can reveal non-obvious effects of cell density, proliferation dynamics and spatial constraints on the effectiveness of control actions (Fiore et al., 2017).

DISCUSSION

We have shown how multi-agent models can be applied to many areas of synthetic biology. The core features of these models provide insight into some of the basic building blocks and mechanisms needed for collective behaviors to emerge and, we believe, may offer a means to support the future predictive design of collective behaviors.

A major hurdle to the widespread use of multi-agent modeling is the need to define and simulate complex models (Grimm et al., 2006). Although computational frameworks have been available since the 1980s to support this process, it is only during the past decade that tools have been

tailored for synthetic biology applications and reached sufficient performance (Gorochowski et al., 2012; Oishi and Klavins, 2014; Goñi-Moreno and Amos, 2015). More recently, the effective use of highly parallel computing resources has expanded the complexity of biological models that can be simulated (Rudge et al., 2012; Naylor et al., 2017; Li et al., 2019; Cooper et al., 2020). Automated coarse-graining of representations enable faster simulation without impacting on the accuracy of predictions (Graham et al., 2017), while advanced tools allow verification, validation and uncertainty quantification for such simulations (Richardson et al., 2020).

Improved simulations do not only speed up the time to an answer but may open up opportunities to create new types of computational design environments. For example, high-performance models coupled to virtual reality allow for multiple researchers to interactively manipulate a system and immediately observe the outcomes of their design decisions. Such capabilities have already begun to be used for molecular design (O'Connor et al., 2018) and when coupled to machine learning, offer a unique setting in which to explore complex high-dimensional datasets that are common in biology. They also allow for essential features to be distilled that can then be used to guide predictive design. Furthermore, hybrid approaches become possible where computational models dynamically augment an experimental setup by controlling physical features such as light (Rubio Denniss et al., 2019) or magnetism (Carlsen et al., 2014). If agents within the experimental system are responsive to these stimuli, then various forms of interaction can be externally programmed and rapidly explored to better understand the necessary conditions for a particular collective behavior to emerge. Once a desired set of rules for the interactions is found, the agents can be modified to implement these autonomously, removing the need for external control.

As synthetic biology moves beyond simple parts and circuits, and toward large-scale/multicellular systems, the available repertoire of design tools must also expand to support new requirements. Multi-agent modeling is perfectly placed to help make this leap and usher in new biological design methods focused on the engineering of emergent collective behaviors. Not only will this allow functionalities to span length scales, but it will also provide a way to engineer across the organizational levels of life through hierarchical composition of multi-scale models, from basic molecules and cells through to entire ecosystems.

AUTHOR CONTRIBUTIONS

TG, SH, J-UK, LM, NS, and T-YT wrote the manuscript. All other authors helped with editing and provided the feedback.

FUNDING

This work captured discussions between participants at the "Multi-agent modeling meets synthetic biology" workshop held on the 16–17 May 2019 at the University of Bristol, UK and funded by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (Grant No. BB/L01386X/1). TG was supported by a Royal Society University Research Fellowship (Grant No. UF160357). DD and VB were supported by the University of Bristol and the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology (Grant No. EP/L016494/1). LB was supported by EPSRC (Grant No. EP/P017134/1-CONDSYC). AF received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (Grant No. 770835). LM was supported by the Medical Research Council (Grant No. MR/N021444/1), and the Engineering and Physical Sciences Research Council (Grant Nos. EP/R041695/1 and EP/S01876X/1). SM was supported by a Mexico Consejo Nacional de Ciencia y Tecnología (CONACYT) Ph.D. scholarship. TF and J-UK are grateful to the UK National Centre for the Replacement, Refinement & Reduction of Animals in Research (NC3Rs) for funding their development of individual-based models (IBMs) for the gut environment (eGUT Grant No. NC/K000683/1 and Ph.D. training Grant No. NC/R001707/1). SH, NS, and SM received funding from the European Union's Horizon 2020 FET Open programme (Grant No. 800983). T-YT acknowledged financial support from the MaxSynBio Consortium (jointly funded by the Federal Ministry of Education and Research, Germany and the Max Planck Society) and the MPI—CBG and the Cluster of Excellence Physics of Life of TU Dresden and EXC-1056 for funding.

REFERENCES

Adamala, K. P., Martin-Alarcon, D. A., Guthrie-Honea, K. R., and Boyden, E. S. (2017). Engineering genetic circuit interactions within and between synthetic minimal cells. *Nat. Chem.* 9, 431–439. doi: 10.1038/nchem.2644

Alvarez, M. M., Guzmán, A., and Elías, M. (2005). Experimental visualization of mixing pathologies in laminar stirred tank bioreactors. *Chem. Eng. Sci.* 60, 2449–2457. doi: 10.1016/j.ces.2004.11.049

Aoki, S. K., Lillacci, G., Gupta, A., Baumschlager, A., Schweingruber, D., and Khammash, M. (2019). A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* 570, 533–537. doi: 10.1038/s41586-019-1321-1

Ashby, W. R. (1991). "Principles of the self-organizing system," in *Facets of Systems Science*, ed. G. J. Klir (Boston, MA: Springer US), 521–536. doi: 10.1007/978-1-4899-0718-9_38

Ausländer, S., and Fussenegger, M. (2016). Engineering gene circuits for mammalian cell–based applications. *Cold Spring Harb. Perspect. Biol.* 8:a023895. doi: 10.1101/cshperspect.a023895

Bacchus, W., Lang, M., El-Baba, M. D., Weber, W., Stelling, J., and Fussenegger, M. (2012). Synthetic two-way communication between mammalian cells. *Nat. Biotechnol.* 30, 991–996. doi: 10.1038/nbt.2351

Bartoli, V., Meaker, G. A., di Bernardo, M., and Gorochowski, T. E. (2020). Tunable genetic devices through simultaneous control of transcription and translation. *Nat. Commun.* 11:2095. doi: 10.1038/s41467-020-15653-7

Bayley, H., Cronin, B., Heron, A., Holden, M. A., Hwang, W. L., Syeda, R., et al. (2008). Droplet interface bilayers. *Mol. Biosyst.* 4, 1191–1208. doi: 10.1039/B808893D

Ben Said, S., and Or, D. (2017). Synthetic microbial ecology: engineering habitats for modular consortia. *Front. Microbiol.* 8:1125. doi: 10.3389/fmicb.2017.01125

Caffrey, J. R., Hughes, B. D., Britto, J. M., and Landman, K. A. (2014). An in silico agent-based model demonstrates Reelin function in directing lamination of neurons during cortical development. *PLoS One* 9:e110415. doi: 10.1371/journal.pone.0110415

Carlsen, R. W., Edwards, M. R., Zhuang, J., Pacoret, C., and Sitti, M. (2014). Magnetic steering control of multi-cellular bio-hybrid microswimmers. *Lab. Chip* 14, 3850–3859. doi: 10.1039/C4LC00707G

Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., et al. (2018). Burden-driven feedback control of gene expression. *Nat. Methods* 15, 387–393. doi: 10.1038/nmeth.4635

Chen, I. A., Roberts, R. W., and Szostak, J. W. (2004). The emergence of competition between model protocells. *Science* 305:1474. doi: 10.1126/science.1100757

Conde-Pueyo, N., Vidiella, B., Sardanyés, J., Berdugo, M., Maestre, T. F., de Lorenzo, V., et al. (2020). Synthetic biology for terraformation lessons from mars, earth, and the microbiome. *Life* 10:14. doi: 10.3390/life10020014

Cooper, F. R., Baker, R. E., Bernabeu, M. O., Bordas, R., Bowler, L., Bueno-Orovio, A., et al. (2020). Chaste: cancer, heart and soft tissue environment. *J. Open Source Softw.* 5:1848. doi: 10.21105/joss.01848

Deisboeck, T. S., and Couzin, I. D. (2009). Collective behavior in cancer cell populations. *BioEssays* 31, 190–197. doi: 10.1002/bies.200800084

Din, M. O., Danino, T., Prindle, A., Skalak, M., Selimkhanov, J., Allen, K., et al. (2016). Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* 536, 81–85. doi: 10.1038/nature18930

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A. G. (2014). Defining an essential transcription factor program for naïve pluripotency. *Science* 344:1156. doi: 10.1126/science.1248882

Ellner, S. P., and Guckenheimer, J. (2011). *Dynamic Models in Biology*. Princeton, NJ: Princeton University Press.

Fedorec, A. J. H., Karkaria, B. D., Sulu, M., and Barnes, C. P. (2019). Killing in response to competition stabilises synthetic microbial consortia. *bioRxiv* [Preprint]. doi: 10.1101/2019.12.23.887331

Fiore, G., Matyjaszkiewicz, A., Annunziata, F., Grierson, C., Savery, N. J., Marucci, L., et al. (2017). In-Silico analysis and implementation of a multicellular feedback control strategy in a synthetic bacterial consortium. *ACS Synth. Biol.* 6, 507–517. doi: 10.1021/acssynbio.6b00220

Gallaher, J. A., Enriquez-Navas, P. M., Luddy, K. A., Gatenby, R. A., and Anderson, A. R. A. (2018). Spatial heterogeneity and evolutionary dynamics modulate time to recurrence in continuous and adaptive cancer therapies. *Cancer Res.* 78:2127. doi: 10.1158/0008-5472.CAN-17-2649

Glen, C. M., Kemp, M. L., and Voit, E. O. (2019). Agent-based modeling of morphogenetic systems: advantages and challenges. *PLoS Comput. Biol.* 15:e1006577. doi: 10.1371/journal.pcbi.1006577

Godwin, S., Ward, D., Pedone, E., Homer, M., Fletcher, A. G., and Marucci, L. (2017). An extended model for culture-dependent heterogenous gene expression and proliferation dynamics in mouse embryonic stem cells. *Npj Syst. Biol. Appl.* 3:19. doi: 10.1038/s41540-017-0020-5

Goldstein, J. (1999). Emergence as a construct: history and issues. *Emergence* 1, 49–72. doi: 10.1207/s15327000em0101_4

Goñi-Moreno, A., and Amos, M. (2015). "DiSCUS: a simulation platform for conjugation computing," in *Unconventional Computation and Natural Computation*, eds C. S. Calude and M. J. Dinneen (Cham: Springer International Publishing), 181–191. doi: 10.1007/978-3-319-21819-9_13

González-Cabaleiro, R., Mitchell, A. M., Smith, W., Wipat, A., and Ofiţeru, I. D. (2017). Heterogeneity in pure microbial systems: experimental measurements and modeling. *Front. Microbiol.* 8:1813. doi: 10.3389/fmicb.2017.01813

Gorochowski, T. E. (2016). Agent-based modelling in synthetic biology. *Essays Biochem.* 60:325. doi: 10.1042/EBC20160037

Gorochowski, T. E., Espah Borujeni, A., Park, Y., Nielsen, A. A., Zhang, J., Der, B. S., et al. (2017). Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol.* 13:952. doi: 10.15252/msb.20167461

Gorochowski, T. E., Matyjaszkiewicz, A., Todd, T., Oak, N., Kowalska, K., Reid, S., et al. (2012). BSim: an agent-based tool for modeling bacterial populations in systems and synthetic biology. *PLoS One* 7:e42790. doi: 10.1371/journal.pone.0042790

Gorochowski, T. E., and Richardson, T. O. (2017). "How behaviour and the environment influence transmission in mobile groups," in *Temporal Network Epidemiology*, eds N. Masuda and P. Holme (Singapore: Springer), 17–42. doi: 10.1007/978-981-10-5287-3_2

Graham, J. A., Essex, J. W., and Khalid, S. (2017). PyCGTOOL: automated generation of coarse-grained molecular dynamics models from atomistic trajectories. *J. Chem. Inf. Model.* 57, 650–656. doi: 10.1021/acs.jcim.7b00096

Greco, F. V., Tarnowski, M. J., and Gorochowski, T. E. (2019). Living computers powered by biochemistry. *Biochemist* 41, 14–18. doi: 10.1042/bio04103014

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecol. Model.* 198, 115–126. doi: 10.1016/j.ecolmodel.2006.04.023

Grozinger, L., Amos, M., Gorochowski, T. E., Carbonell, P., Oyarzún, D. A., Stoof, R., et al. (2019). Pathways to cellular supremacy in biocomputing. *Nat. Commun.* 10:5250. doi: 10.1038/s41467-019-13232-z

Haringa, C., Tang, W., Wang, G., Deshmukh, A. T., van Winden, W. A., Chu, J., et al. (2018). Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: towards rational scale-down and design optimization. *Chem. Eng. Sci.* 175, 12–24. doi: 10.1016/j.ces.2017.09.020

Hasty, J., McMillen, D., and Collins, J. J. (2002). Engineered gene circuits. *Nature* 420, 224–230. doi: 10.1038/nature01257

Hauert, S., Berman, S., Nagpal, R., and Bhatia, S. N. (2013). A computational framework for identifying design guidelines to increase the penetration of targeted nanoparticles into tumors. *Nano Today* 8, 566–576. doi: 10.1016/j.nantod.2013.11.001

Hauert, S., and Bhatia, S. N. (2014). Mechanisms of cooperation in cancer nanomedicine: towards systems nanotechnology. *Trends Biotechnol.* 32, 448–455. doi: 10.1016/j.tibtech.2014.06.010

Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M., and Kreft, J.-U. (2016). Advancing microbial sciences by individual-based modelling. *Nat. Rev. Microbiol.* 14, 461–471. doi: 10.1038/nrmicro.2016.62

Hsiao, V., Swaminathan, A., and Murray, R. M. (2018). Control theory for synthetic biology: recent advances in system characterization, control design, and controller implementation for synthetic biology. *IEEE Control Syst. Mag.* 38, 32–62. doi: 10.1109/mcs.2018.2810459

Jacob, C., Litorco, J., and Lee, L. (2004). "Immunity through swarms: agent-based simulations of the human immune system," in *Artificial Immune Systems*, eds G. Nicosia, V. Cutello, P. J. Bentley, and J. Timmis (Berlin: Springer), 400–412. doi: 10.1007/978-3-540-30220-9_32

Jiang, X., Zerfaß, C., Feng, S., Eichmann, R., Asally, M., Schäfer, P., et al. (2018). Impact of spatial organization on a novel auxotrophic interaction among soil microbes. *ISME J.* 12, 1443–1456. doi: 10.1038/s41396-018-0095-z

Joesaar, A., Yang, S., Bögels, B., van der Linden, A., Pieters, P., Kumar, B. V. V. S. P., et al. (2019). DNA-based communication in populations of synthetic protocells. *Nat. Nanotechnol.* 14, 369–378. doi: 10.1038/s41565-019-0399-9

Kazamia, E., Aldridge, D. C., and Smith, A. G. (2012). Synthetic ecology – A way forward for sustainable algal biofuel production? *Photosynth. Microorg. Bio Fuel Prod. Sun Light* 162, 163–169. doi: 10.1016/j.jbiotec.2012.03.022

Kempes, C. P., Koehl, M. A. R., and West, G. B. (2019). The scales that limit: the physical boundaries of evolution. *Front. Ecol. Evol.* 7:242. doi: 10.3389/fevo.2019.00242

Kerényi, Á, Bihary, D., Venturi, V., and Pongor, S. (2013). Stability of multispecies bacterial communities: signaling networks may stabilize microbiomes. *PLoS One* 8:e57947. doi: 10.1371/journal.pone.0057947

Kreft, J.-U., Griffin, B. M., and González-Cabaleiro, R. (2020). Evolutionary causes and consequences of metabolic division of labour: why anaerobes do and aerobes don't. *Curr. Opin. Biotechnol.* 62, 80–87. doi: 10.1016/j.copbio.2019.08.008

Kreft, J.-U., Plugge, C. M., Prats, C., Leveau, J. H. J., Zhang, W., and Hellweger, F. L. (2017). From genes to ecosystems in microbiology: modeling approaches and the importance of individuality. *Front. Microbiol.* 8:2299. doi: 10.3389/fmicb.2017.02299

Kumar, M., Ji, B., Zengler, K., and Nielsen, J. (2019). Modelling approaches for studying the microbiome. *Nat. Microbiol.* 4, 1253–1267. doi: 10.1038/s41564-019-0491-9

Kuznetsov, Y. A. (2004). *Elements of Applied Bifurcation Theory*. New York, NY: Springer.

Lambert, B., MacLean, A. L., Fletcher, A. G., Combes, A. N., Little, M. H., and Byrne, H. M. (2018). Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *J. Math. Biol.* 76, 1673–1697. doi: 10.1007/s00285-018-1208-z

Lee, S. Y., and Kim, H. U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 33, 1061–1072. doi: 10.1038/nbt.3365

Lentini, R., Santero, S. P., Chizzolini, F., Cecchi, D., Fontana, J., Marchioretto, M., et al. (2014). Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour. *Nat. Commun.* 5:4012. doi: 10.1038/ncomms5012

Li, B., Taniguchi, D., Gedara, J. P., Gogulancea, V., Gonzalez-Cabaleiro, R., Chen, J., et al. (2019). NUFEB: a massively parallel simulator for individual-based modelling of microbial communities. *PLoS Comput. Biol.* 15:e1007125. doi: 10.1371/journal.pcbi.1007125

Li, M., Huang, X., Tang, T.-Y. D., and Mann, S. (2014). Synthetic cellularity based on non-lipid micro-compartments and protocell models. *Synth. Biol. Synth. Biomol.* 22, 1–11. doi: 10.1016/j.cbpa.2014.05.018

Louca, S., and Doebeli, M. (2016). Transient dynamics of competitive exclusion in microbial communities. *Environ. Microbiol.* 18, 1863–1874. doi: 10.1111/1462-2920.13058

McCardell, R. D., Huang, S., Green, L. N., and Murray, R. M. (2017). Control of bacterial population density with population feedback and molecular sequestration. *bioRxiv* [Preprint]. doi: 10.1101/225045

McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., et al. (2004). Fermented beverages of pre- and proto-historic China. *Proc. Natl. Acad. Sci. U.S.A.* 101:17593. doi: 10.1073/pnas.0407921102

Menolascina, F., Fiore, G., Orabona, E., De Stefano, L., Ferry, M., Hasty, J., et al. (2014). In-Vivo real-time control of protein expression from endogenous and synthetic gene networks. *PLoS Comput. Biol.* 10:e1003625. doi: 10.1371/journal.pcbi.1003625

Metzcar, J., Wang, Y., Heiland, R., and Macklin, P. (2019). A review of cell-based computational modeling in cancer biology. *JCO Clin. Cancer Inform.* 3, 1–13. doi: 10.1200/CCI.18.00069

Montes-Olivas, S., Marucci, L., and Homer, M. (2019). Mathematical models of organoid cultures. *Front. Genet.* 10:873. doi: 10.3389/fgene.2019.00873

Morsut, L., Roybal, K. T., Xiong, X., Gordley, R. M., Coyle, S. M., Thomson, M., et al. (2016). Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell* 164, 780–791. doi: 10.1016/j.cell.2016.01.012

Naylor, J., Fellermann, H., Ding, Y., Mohammed, W. K., Jakubovics, N. S., Mukherjee, J., et al. (2017). Simbiotics: a multiscale integrative platform for 3D modeling of bacterial populations. *ACS Synth. Biol.* 6, 1194–1210. doi: 10.1021/acssynbio.6b00315

Nicolis, G., and Prigogine, I. (1989). *Exploring Complexity: An Introduction*. New York, NY: W.H. Freeman and Company.

Niederholtmeyer, H., Sun, Z. Z., Hori, Y., Yeung, E., Verpoorte, A., Murray, R. M., et al. (2015). Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* 4:e09771. doi: 10.7554/eLife.09771

Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., et al. (2016). Genetic circuit design automation. *Science* 352:aac7341. doi: 10.1126/science.aac7341

Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004

Nieß, A., Löffler, M., Simen, J. D., and Takors, R. (2017). Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of *E. coli* cultures in large-scale bioreactors: experimental evidence and mathematical model. *Front. Microbiol.* 8:1195. doi: 10.3389/fmicb.2017.01195

O'Connor, M., Deeks, H. M., Dawn, E., Metatla, O., Roudaut, A., Sutton, M., et al. (2018). Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Sci. Adv.* 4:eaat2731. doi: 10.1126/sciadv.aat2731

Oishi, K., and Klavins, E. (2014). Framework for engineering finite state machines in gene regulatory networks. *ACS Synth. Biol.* 3, 652–665. doi: 10.1021/sb4001799

Pedone, E., Postiglione, L., Aulicino, F., Rocca, D. L., Montes-Olivas, S., Khazim, M., et al. (2019). A tunable dual-input system for on-demand dynamic gene expression regulation. *Nat. Commun.* 10:4481. doi: 10.1038/s41467-019-12329-9

Perez-Carrasco, R., Guerrero, P., Briscoe, J., and Page, K. M. (2016). Intrinsic noise profoundly alters the dynamics and steady state of morphogen-controlled bistable genetic switches. *PLoS Comput. Biol.* 12:e1005154. doi: 10.1371/journal.pcbi.1005154

Perfahl, H., Hughes, B. D., Alarcón, T., Maini, P. K., Lloyd, M. C., Reuss, M., et al. (2017). 3D hybrid modelling of vascular network formation. *J. Theor. Biol.* 414, 254–268. doi: 10.1016/j.jtbi.2016.11.013

Prigogine, I., and Nicolis, G. (1985). "Self-organisation in nonequilibrium systems: towards a dynamics of complexity," in *Bifurcation Analysis: Principles, Applications and Synthesis*, eds M. Hazewinkel, R. Jurkovich, and J. H. P. Paelinck (Dordrecht: Springer Netherlands), 3–12. doi: 10.1007/978-94-009-6239-2_1

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., et al. (2013). lessons learned from quantitative dynamical modeling in systems biology. *PLoS One* 8:e74335. doi: 10.1371/journal.pone.0074335

Richardson, R. A., Wright, D. W., Edeling, W., Jancauskas, V., Lakhlili, J., and Coveney, P. V. (2020). EasyVVUQ: a library for verification, validation and uncertainty quantification in high performance computing. *J. Open Res. Softw.* 8:11. doi: 10.5334/jors.303

Rovinskii, A. B., and Zhabotinskii, A. M. (1984). Mechanism and mathematical model of the oscillating bromate-ferroin-bromomalonic acid reaction. *J. Phys. Chem.* 88, 6081–6084. doi: 10.1021/j150669a001

Rubio Denniss, A. M., Gorochowski, T. E., and Hauert, S. (2019). "Augmented reality for the engineering of collective behaviours in microsystems," in *Proceedings of the 2019 International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, Helsinki.

Rudge, T. J., Steiner, P. J., Phillips, A., and Haseloff, J. (2012). Computational modeling of synthetic microbial biofilms. *ACS Synth. Biol.* 1, 345–352. doi: 10.1021/sb300031n

Schuerle, S., Soleimany, A. P., Yeh, T., Anand, G. M., Häberli, M., Fleming, H. E., et al. (2019). Synthetic and living micropropellers for convection-enhanced nanoparticle transport. *Sci. Adv.* 5:eaav4803. doi: 10.1126/sciadv.aav4803

Schuster, S., Kreft, J.-U., Brenner, N., Wessely, F., Theißen, G., Ruppin, E., et al. (2010). Cooperation and cheating in microbial exoenzyme production – Theoretical analysis for biotechnological applications. *Biotechnol. J.* 5, 751–758. doi: 10.1002/biot.200900303

Scott, J. G., Fletcher, A. G., Anderson, A. R. A., and Maini, P. K. (2016). Spatial metrics of tumour vascular organisation predict radiation efficacy in a computational model. *PLoS Comput. Biol.* 12:e1004712. doi: 10.1371/journal.pcbi.1004712

Scott, J. G., Maini, P. K., Anderson, A. R., and Fletcher, A. G. (2019). Inferring tumor proliferative organization from phylogenetic tree measures in a computational model. *Syst. Biol.* syz070. doi: 10.1093/sysbio/syz070

Shou, W., Ram, S., and Vilar, J. M. G. (2007). Synthetic cooperation in engineered yeast populations. *Proc. Natl. Acad. Sci. U.S.A.* 104:1877. doi: 10.1073/pnas.0610575104

Shreiner, A. B., Kao, J. Y., and Young, V. B. (2015). The gut microbiome in health and in disease. *Curr. Opin. Gastroenterol.* 31, 69–75.

Solé, R. V., Montañez, R., and Duran-Nebreda, S. (2015). Synthetic circuit designs for earth terraformation. *Biol. Direct.* 10:37. doi: 10.1186/s13062-015-0064-7

Stillman, N. R., Kovacevic, M., Balaz, I., and Hauert, S. (2020). In silico modelling of cancer nanomedicine, across scales and transport barriers. *Npj Comput. Mater.* (in press).

Stopka, A., Kokic, M., and Iber, D. (2019). Cell-based simulations of biased epithelial lung growth. *Phys. Biol.* 17:016006. doi: 10.1088/1478-3975/ab5613

Tang, T.-Y. D., Cecchi, D., Fracasso, G., Accardi, D., Coutable-Pennarun, A., Mansy, S. S., et al. (2018). Gene-mediated chemical communication in synthetic protocell communities. *ACS Synth. Biol.* 7, 339–346. doi: 10.1021/acssynbio.7b00306

Taroni, A. (2015). 90 years of the Ising model. *Nat. Phys.* 11, 997–997. doi: 10.1038/nphys3595

Toettcher, J. E., Gong, D., Lim, W. A., and Weiner, O. D. (2011). Light-based feedback for controlling intracellular signaling dynamics. *Nat. Methods* 8, 837–839. doi: 10.1038/nmeth.1700

Treloar, N. J., Fedorec, A. J. H., Ingalls, B., and Barnes, C. P. (2020). Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLoS Comput. Biol.* 16:e1007783. doi: 10.1371/journal.pcbi.1007783

Velazquez, J. J., Su, E., Cahan, P., and Ebrahimkhani, M. R. (2018). Programming morphogenesis through systems and synthetic biology. *Trends Biotechnol.* 36, 415–429. doi: 10.1016/j.tibtech.2017.11.003

Waclaw, B., Bozic, I., Pittman, M. E., Hruban, R. H., Vogelstein, B., and Nowak, M. A. (2015). A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* 525, 261–264. doi: 10.1038/nature14971

Ward, D., Montes Olivas, S., Fletcher, A., Homer, M., and Marucci, L. (2020). Cross-talk between Hippo and Wnt signalling pathways in intestinal crypts: insights from an agent-based model. *Comput. Struct. Biotechnol. J.* 18, 230–240. doi: 10.1016/j.csbj.2019.12.015

West, G. B., Woodruff, W. H., and Brown, J. H. (2002). Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl. Acad. Sci. U.S.A.* 99:2473. doi: 10.1073/pnas.012579799

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* 10, 2557–2568. doi: 10.1038/ismej.2016.45

Wu, S., Liu, J., Liu, C., Yang, A., and Qiao, J. (2020). Quorum sensing for population-level control of bacteria and potential therapeutic applications. *Cell. Mol. Life Sci.* 77, 1319–1343. doi: 10.1007/s00018-019-03326-8

Zadorin, A. S., Rondelez, Y., Gines, G., Dilhas, V., Urtel, G., Zambrano, A., et al. (2017). Synthesis and materialization of a reaction–diffusion French flag pattern. *Nat. Chem.* 9, 990–996. doi: 10.1038/nchem.2770

Zerfaß, C., Chen, J., and Soyer, O. S. (2018). Engineering microbial communities using thermodynamic principles and electrical interfaces. *Energy Biotechnol. Environ. Biotechnol.* 50, 121–127. doi: 10.1016/j.copbio.2017.12.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gorochowski, Hauert, Kreft, Marucci, Stillman, Tang, Bandiera, Bartoli, Dixon, Fedorec, Fellermann, Fletcher, Foster, Giuggioli, Matyjaszkiewicz, McCormick, Montes Olivas, Naylor, Rubio Denniss and Ward. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.