

Document Version

Final published version

Licence

CC BY

Citation (APA)

Bauer, M., Goyal, A., Goyal, S., Reddy, G., Chakrabarti, S., Desai, M. M., Gilpin, W., Grilli, J., Jain, S., & More Authors (2026). Unifying theories in high-dimensional biophysics: approaches, challenges and opportunities. *npj systems biology and applications*, 12(1), Article 43. <https://doi.org/10.1038/s41540-026-00680-9>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Unifying theories in high-dimensional biophysics: approaches, challenges and opportunities

Marianne Bauer, Akshit Goyal, Sidhartha Goyal, Gautam Reddy, Shaon Chakrabarti, Michael M. Desai, William Gilpin, Jacopo Grilli, Kabir Husain, Sanjay Jain, Mohit Kumar Jolly, Kyogo Kawaguchi, Aneta Koseska, Milo Lin, Leelavati Narlikar, Simone Pigolotti, Archishman Raju, Krishna Shrinivas, Rahul Siddharthan, Greg J. Stephens, Andreas Tiffeau-Mayer & Suriyanarayanan Vaikuntanathan



Across biological subdisciplines, the last decade has seen an explosion of high-dimensional datasets. At the ICTS workshop ‘Unifying Theories in High-Dimensional Biophysics’, we discussed whether this high dimensionality poses a challenge or an opportunity for theoretically describing, understanding and predicting biological systems. We discussed methods, models and frameworks that can be used for this purpose. This Comment summarizes our discussions from the perspectives of individual participants.

The high-dimensional nature of biological systems creates a formidable challenge in making conceptual advances. The vast amount of datasets across diverse fields, ranging from development, immunology, ecology, evolution, neuroscience, and behavior^{1–10}, offers an opportunity to glean quantitative insights. Nevertheless, we lack common theoretical frameworks to extract generalizable principles.

One central idea emerging from recent work is that low-dimensional models can be surprisingly good at explaining empirical observations while providing specific testable predictions. Examples include the use of sloppy models to create simplifying descriptions of biochemical networks^{11–13}, coarse-grained resource allocation models in microbial metabolism^{14,15}, low-dimensional dynamical systems models of development^{16,17}, or low-dimensional structure in genotype-phenotype maps^{18,19}, amongst others. Machine learning helps find some of this low-dimensional structure, for example, by using sequence data from developing organisms²⁰, tumors and immune repertoires²¹, and from large-scale neural recordings^{22,23}. However, different subfields have developed different analytical and machine learning approaches for tackling high-dimensional complexity.

In this brief interdisciplinary viewpoint, we share thoughts discussed in the realm of high-dimensional biophysics at a workshop at the International Centre for Theoretical Sciences (ICTS) in the summer of 2025. Is there a common mathematical language that can be used to describe emergent simplicity across systems? Are there common biological principles, such as evolvability and functional robustness, that make low-dimensional descriptions possible? How do we understand if a system truly operates in

high dimensions, for example, the fine-scale diversity in microbial ecosystems? What are good ‘null’ models, and what details matter for explaining phenomenology?

We gathered contributions from theoretical and computational physicists at our workshop working in diverse areas of biology, including development, ecology and evolution, immunology, neuroscience, and protein science. These contributions highlight commonalities, differences, and challenges of high-dimensional biology and identify opportunities for progress toward a unifying framework. We structured individual contributions by scale of their dominant application topics. We note that the references in this Comment are not meant to be complete (or exhaustive), but rather indicative of ideas and approaches that came to the forefront during the meeting.

Molecules and proteins

Kabir Husain. A physicist once interrupted a talk I was giving on protein chaperones – molecules that assist in protein folding – with a question—asked with a certain degree of incredulity. How, they wanted to know, could a single chaperone help so many different proteins fold? Did it somehow memorize the correct folded configuration of each one?

Indeed, recognizing that biology operates in what we might call a “high-dimensional” space raises unsettling questions of this kind. One defining feature of a high-dimensional space is how easy it is to get lost within it: while random diffusion might suffice to find a target in one or two dimensions, the likelihood drops precipitously as the number of dimensions increases. This would seem to suggest that robust, reliable systems should be low-dimensional, with only a few moving parts. Yet biology appears to defy this intuition. It is full of diverse, intricately structured entities that assemble precisely and reproducibly—whether it be the sequence of nucleotides in a genome, the arrangement of atoms in a folded protein, or the patterning of cells in a developing organism. How does that work?

Questions of this breadth inevitably have many system-specific answers. But some overarching strategies can be discerned. First, the “correct” or functional biological configuration often corresponds to one that is energetically most favorable—as in protein folding—or that has the longest residence time, as in immune recognition. Second, biological search processes are rarely random walks. They follow stereotyped trajectories through energy landscapes, guided by structures such as folding funnels or Waddington’s developmental “landscape.” Third—and perhaps least generalizable—when errors occur, biology often has ways of detecting them without needing to know exactly what went wrong. For instance, some chaperones

recognize exposed hydrophobic patches that would normally be buried within a protein's core, while DNA repair mechanisms detect distortions in the regular shape of Watson-Crick base pairs. In short, the physical architecture of biological systems seems remarkably well adapted to navigating high-dimensional spaces reliably.

What do we gain by thinking at this level of generality? Beyond the intellectual pleasure of tracing analogies across systems, such reflections suggest constraints on what kinds of physical systems can support something we would recognize as “biology.” If evolution—another search through an immense, high-dimensional space—were to begin again on a different world, many details would surely differ. However, I suspect that the basic challenge of finding one's way through high-dimensional spaces would impose severe constraints on the physical make-up of any living system.

Milo Lin. Proteins are chains of hundreds to thousands of amino acids that self-fold into their functional configurations. Because proteins carry out most of the biomolecular processes responsible for life, how the amino acid sequence encodes function is a high-dimensional question upon which many other questions, at larger scales, are built. Yet this question is largely unresolved for two reasons. First, although much progress has been made in predicting protein structure from sequence, these advances are black boxes that have not yielded unifying principles of protein folding. Second, the assumption that structure (i.e., protein folding) is sufficient to prescribe function is incomplete. For enzymatic proteins, the activity of the active site is often coupled to the dynamics of the rest of the protein. Many proteins are allosteric; allostery means that these proteins can modulate the activity of their active site depending on the detection of signals at other sites of the protein. This allows proteins to be input-output devices, which is crucial for their role as the building blocks of biological circuits. In many proteins, allostery depends on correlated dynamics. We have found that treating proteins as coupled networks of conformational fluctuations enables quantitative prediction of protein allostery. At the level of amino acids, protein fluctuations can be well-approximated as digital transformations between discrete states, allowing us to borrow approaches from spin systems and glasses. This meeting introduced connections of protein function with other fields, in particular ecological networks, and I anticipate future progress to be advanced by the cross-application of ideas and approaches between these fields operating at very different scales.

Suriyanarayanan Vaikuntanathan. High-dimensionality poses a fundamental challenge in deciphering complex biological codes like the glycan code, where the combinatorial diversity of glycan structures creates a potentially overwhelming information space that cells must somehow read and interpret. Our recent work in *Nature Communications*²⁴ demonstrates how principles from non-equilibrium thermodynamics offer a promising way forward: molecular promiscuity in receptor-glycan interactions can actually serve as a feature rather than a bug, effectively reducing the dimensionality required for accurate biological readout. This insight connects naturally to techniques from computer science, particularly compressed sensing, which shows how high-dimensional signals can be recovered from surprisingly few measurements when the underlying structure is sparse or compressible. More broadly, I see potential for bidirectional exchange between fields—while computational frameworks like compressed sensing illuminate how biological systems might efficiently decode complex molecular information, the thermodynamic principles governing promiscuous binding and selective readout in biology could inspire novel computational paradigms that go beyond conventional compressed sensing approaches. These bio-inspired algorithms might

offer new ways to handle high-dimensional data in artificial systems, particularly where noise, dynamics, and energy constraints mirror those faced by living cells.

The nucleus: chromatin and gene expression

Kyogo Kawaguchi. Over the past decades, biology has entered a high-dimensional era, with massive data collection across gene expression dynamics, chromatin profiles, neuronal activity, and more. This expansion partly realizes the vision of nonlinear dynamics and self-organization theory, which converged into systems biology around the 2000s. With this abundance of measurable quantities, a central challenge now is to ask whether we can already constrain which models are sensible and which are falsifiable.

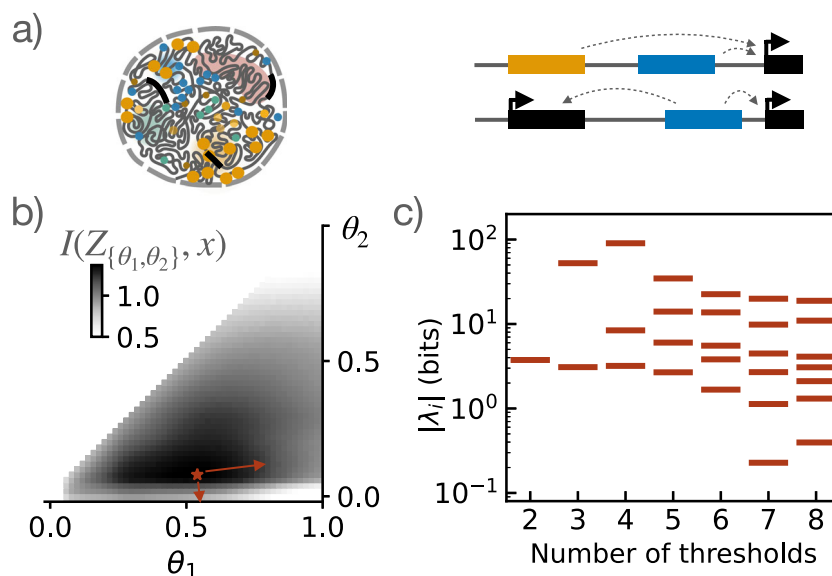
Since biological dynamics are inevitably multiscale and complex, it has been natural to begin from a common framework such as dynamical-systems approaches derived from chemical reaction networks, which have been particularly influential in explaining cell fate decisions. But these models often neglect spatial components—such as phase separation in the nucleus and polymer dynamics in chromatin—that almost certainly contribute to fate determination. Other strategies from physics, such as coarse-graining and the use of phase transitions and macroscopic order parameters, have proven useful for capturing at least parts of those phenomena. Beyond these, heuristic and data-driven approaches are rapidly advancing, highlighting the need for still broader theoretical frameworks that can connect diverse mechanistic models with purely empirical descriptions.

Leelavati Narlikar and Rahul Siddharthan. Modern molecular biology is driven by high-throughput data generation, starting with the advent of new “next-generation” sequencing technologies about 20 years ago. These technologies, apart from sequencing genomes, can be used to perform genome-wide assays for transcript abundance, DNA-protein binding, chromatin accessibility, chromatin interactions, and much more. More recently, single-cell transcriptomics has enabled the generation of transcriptomic and chromatin accessibility data at a single-cell level. The output of such experiments consists of matrices with 1000–100,000 rows (cells) and up to 25,000 columns (genes). Analysis of such inherently high-dimensional data, and interpretation of the results, is a challenge, and the field relies on methods such as UMAP for visualization, which have caused some controversy^{25,26}.

Our talks focused on the specific problems of DNA-protein interaction and transcription initiation, where current datasets are mostly from bulk samples. Even so, genome-wide data exists for 100+ transcription factors and histone modifications, and dozens of cell types, in humans alone. Predictive models for DNA-protein binding, including sequence information, interactions among TFs, and chromatin accessibility data, have been a challenge for many years. We discussed combinatorial binding via mixture models and sequence signatures identified via deep learning. In the near future, single-cell ChIP-seq is expected to become more common, enabling us to learn much more specific modes of binding, but also increasing the challenge of analyzing the data.

At this meeting, we heard from other speakers about the application of dimensionality reduction, deep learning methods, Bayesian methods, information theory, and more, across multiple fields. In particular, multiple talks discussed the use of latent dimensions – dimensions that correspond to hidden variables that cannot be directly observed – to compress data and better understand it. This is something we also spoke about in our mixture models of regulatory data. We expect that such techniques will be developed synchronously across many areas of modern biology, with ample cross-pollination of ideas.

Fig. 1 | Sloppy landscapes of optimal gene regulation. **a** Genes are expressed in the spatially complex geometry of the nucleus, based on multiple genomic regions upstream and/or downstream of the gene that can regulate a single gene together as well as multiple genes. **b** Even in the simplest possible model where these regulatory regions activate expression based on threshold concentrations θ of an input, the optimal positioning of these thresholds that maximizes transfer of developmental information $I(Z_{\{\theta_1, \theta_2\}}; x)$, here for two thresholds, is ‘sloppy’: **c** This implies that the eigenvalues λ of the Hessian matrix at the optimum span several decades (adapted from ref. 31). This spectrum implies that there are some sloppy (eigen-)directions along which the landscape falls off slowly from the optimal value; parameter combinations along this sloppy direction can therefore change significantly while the system remains essentially optimal, suggesting that a diverse set of implementations of regulatory regions can operate near optimality.



Marianne Bauer. Genes are regulated by a myriad of processes, involving the binding of transcription or pioneer factors to one or multiple enhancer or promoter regions that may require spatial contacts or the establishment of a local transcription factor hub²⁷ (see Fig. 1). This regulation is necessary for biological function, for example, correct development or responses to environmental change. While this functional output can be modeled with intuition on participating molecules and mechanism^{28,29}, an alternative approach can be to posit this function explicitly as an optimization goal and to find mechanisms consistent with this optimization^{30–34}. Optimization approaches have been of interest across biological sub-disciplines³⁵, starting with work on neurons^{36,37}, as well as in the context of neural networks, where the loss function or optimization goal is perhaps more but also not always obvious.

Optimization in high dimensions can be difficult, but the high dimensionality of parameters and processes in biological systems provides an advantage to optimization approaches: optima of the loss functions are often ‘sloppy’. The concept of ‘sloppiness’ was introduced in the context of model fitting to describe combinations of parameters that are highly variable without significant loss of fit accuracy¹¹; since the optimization landscape also falls off slowly along specific (sloppy) directions away from the optimum, combinations of parameters may vary significantly without affecting the optimization function. Therefore, for sloppy landscapes, even optimizing a mean performance will bring a solution ‘close’ to an optimum³⁸. That multiple parameter combinations are consistent with near-optimal performance is indeed clear from data-driven theoretical observations - both for optimal regulatory elements³¹, as well as for optimal signals³⁹ - and also directly from experiment: in gene regulation, this observation is consistent with the biological diversity of functional regulatory pathways and constructs.

Even though a challenge of optimization approaches is sometimes the absence of a clear, unifying goal (see, e.g., the discussion in the “Cells and development” section), optimization ideas can present a respectful approach for understanding how biological systems - in gene regulation and beyond - fulfill their function. Nevertheless, it is important to critically assess

the applicability of optimization goals and the importance of constraints in specific systems. Since living systems develop and evolve, investigating optimality also requires taking into account the kinetics or dynamic trajectories of the signal response, an active area of research at the moment^{40–47}. An exciting general direction is to compare the regulatory and developmental dynamics to those of learning in biological and artificial systems, where optimization is more established and where dynamics can help predict outcomes, for example, on whether networks learn or generalize⁴⁸.

Cells and development

Krishna Shrinivas. The living cell is intrinsically high-dimensional. Vast numbers of microscopic parameters, such as molecular concentrations, interactions, and reactions, often tuned by evolution, drive emergent properties of low-dimensional cellular processes. A striking example of this funneling is the organization of the cellular milieu into tens of coexisting compartments called condensates, whose molecular compositions comprise shared and distinct features. In recent years, phase transitions have been identified as central to driving condensate formation. Here, I highlight a few considerations that arise from viewing the cell as a squishy molecular stew rather than a watery, well-mixed soup of molecules.

How do high-dimensional molecular properties channel into emergent condensate features, and which microscopic details—protein sequence, stoichiometry, excluded volume, reactions—actually matter? A related question is how best to encode such microscopic details and what could limit expressivity and capacity in resulting models⁴⁹. In parallel, biological measurements of condensates are typically low-dimensional - visualizable fluorophores on a handful of proteins, or phenotypic readouts such as expression of a downstream gene. From such sparse observables, can we infer higher-dimensional constraints on condensate properties and reveal evolutionary or functional restrictions on these ensembles? Statistical physics suggests possible routes: predictive models that coarse-grain microscopic degrees of freedom, response theories that infer hidden structure from perturbations, and information-geometric approaches that separate stiff from sloppy modes. These considerations resonate with other fields

where distinct microscopic details nonetheless produce rich emergent behavior, offering opportunities for both importing methods as well as identifying unifying principles. Prominent examples discussed at the workshop include microbial ecology, where community-scale functions emerge from diverse consortia with distinct microscopic interactions, and understanding emergence in artificial intelligence, where billions of nodes are distilled into architectures that generalize and solve specific tasks. Ultimately, understanding the ingredients and recipes that underpin cellular stews offers a rich theoretical playground for high-dimensional biology.

Mohit Kumar Jolly. Elucidating the dynamics of high-dimensionality of the regulatory biochemical space at multiple levels - transcriptional, epigenetic, and metabolic - is an open question to understand cell decision-making in development, disease, and reprogramming⁵⁰. We have witnessed a surge in technological advancements that facilitate this information being collated in a dynamic manner experimentally, and the development of new computational methods that can extract at least pseudotime-based projections from static high-dimensional experimental data. Thanks to these advancements, we are inching closer toward mapping the changes in these high-dimensional landscapes that enable or accompany cell-state transitions, thus opening up new avenues to apply many dynamical systems theory approaches to identify which trajectories are allowed or not for cells traverse through this landscape, and how their trajectories can be modified to alter the destined cell-state⁵¹. The concepts of attractor states, frustration, and consequent cell-state canalization are being increasingly used to quantify entropy in gene expression during cell-state transitions, which can eventually allow for tying together the high-dimensional experimental data with low dimensionality of the phenotypic space, using theoretical biophysics-based novel conceptual frameworks⁵²⁻⁵⁷.

Sanjay Jain. Examples of biological systems that exhibit homeostasis or chaos both exist, but do we understand which systems exhibit these two behaviors, and when and why? Homeostasis is the maintenance of the internal state of the system or of a relevant subsystem in response to certain perturbations. An example of homeostasis is the exponential growth culture of a bacterium, where hundreds to thousands of reactions involving similar numbers of molecular species achieve a coordinated steady state within each cell. This non-equilibrium steady state can be understood as a fixed point attractor of a complex dynamical system in a high-dimensional chemical concentration space representing the cell in a fixed environment. Mathematical models of a cell involving the nonlinear chemical reaction kinetics of hundreds of interacting molecular species can be constructed that exhibit this behavior^{58,59}. Stochasticity in the dynamics leads to a stable stationary probability distribution of states localized around the fixed point. This orderly behavior, or homeostasis, is required for the self-reproduction of the bacterium.

On the other hand, in this meeting, we also saw examples of chaotic dynamics in biological systems from Greg Stephens (motion of *C. Elegans*⁴¹, social behavior of Zebrafish, and honey bees). This prompts the question: What is the relationship between the homeostasis necessary for life and the observed chaotic behavior? Can there be a classification of biological systems into those that exhibit stable fixed point or limit cycle type behavior that could be regarded as homeostatic, and those that exhibit chaos? Are there common architectural/circuitry features (feedback, control, or lack of it) that characterize each kind of behavior? Is metabolism homeostatic, and social behavior chaotic? Are eukaryotes (and multicellular organisms) more prone to exhibiting chaotic dynamics compared to prokaryotes (and unicellular organisms)? Is chaotic behavior always built on top of a homeostatic

core? Do the same organisms show both homeostasis and chaos in different subsystems or different environments? Do these behaviors coexist (criticality or “edge of chaos” hypothesis)? Do the hundreds to thousands of interacting chemicals in *E. coli* have a dynamical regime that is chaotic? Maybe it is useful for the bacterium to be homeostatic while the going is good (food available, exponential growth). But when times are not so good (no food, stationary phase), it could be an advantage for the bacterium to be chaotic (more exploratory). It might be interesting to explore this in mathematical models and data.

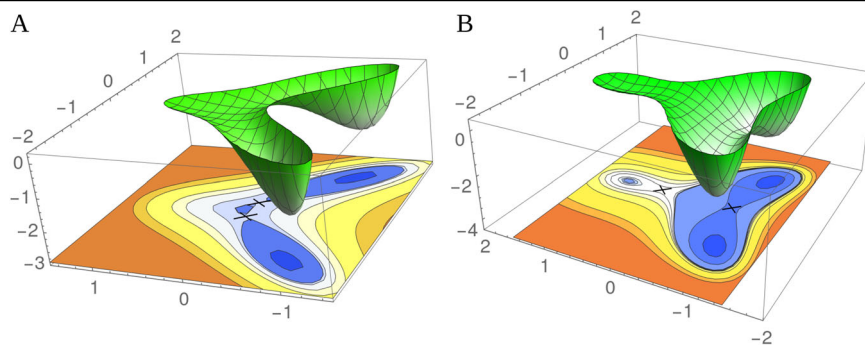
Sid Goyal. Does the effective dimensionality of a biological system change with time or conditions? Our recent dynamical system view of hematopoiesis claimed that cells in the gene expression space display statistical signatures consistent with bifurcations during fate change and that there is a low-dimensional phase plane in gene expression space within which the multistability unfolds⁶⁰. Is this reduced dimensionality only evident during fate transitions? To address this, we asked if we could embed complete cell fate trajectories in low dimensions⁶¹ in a variety of development contexts. In doing so, we inferred a latent regulatory state that controls the dynamics of an individual cell to model multiple lineages. We find that the dimensionality of the low-dimensional space depended on the number of final states, suggesting some limitations on effective dimensionality.

Is this low dimensionality a property of developing systems getting canalized to provide robustness against perturbations? The expectation is that, in contrast to the *in vivo* systems discussed above, differentiated cells, as they reprogram *in vitro*, can, in principle, take many different paths in the high-dimensional space of gene expression. However, we find only two populations with distinct rates of reprogramming⁶². It will be fascinating if these correspond to two different low-dimensional manifolds, or is it an ensemble of different paths that give similar transition rates.

As we accumulate more examples of fate transitions and their dimensionality, it will be fascinating to understand if and how effective dimensionality is controlled and used in biology.

Aneta Koseska. Cells face a fundamental computational problem: they must process time-varying inputs utilizing signaling and genetic networks to make context-dependent decisions and determine cell fate^{39,63,64}. Understanding how the distributed biochemical networks accomplish this requires analyzing the evolving state space of the system, which is inherently high-dimensional. Non-autonomous dynamical systems provide a natural framework to describe the time-dependent trajectories of the network⁶⁵. Our recent work demonstrates that biochemical networks in single cells perform a mapping of classes of dynamic input signals to classes of signaling state-space trajectories, each corresponding to a distinct phenotype⁴⁷. Crucially, this mapping requires the network to possess a form of dynamic memory⁶⁶, which is both necessary and sufficient to reproduce the observed behavior. Even low-dimensional coarse-grained models are capable of capturing this principle, highlighting the minimal mechanistic rules that govern how cells integrate complex inputs and commit to specific fates. At the same time, the high-dimensionality of the biochemical network enables robust separation of signaling trajectories, ensuring a unique mapping to each phenotype. This points to a subtle interplay between the rich structure provided by high-dimensional networks and the minimal features that low-dimensional models capture. Future work should focus on developing principled approaches to uncover how best to extract these minimal representations while preserving the functional benefits of high-dimensional biochemical architectures, providing a unified framework for understanding cellular computation in complex biological systems.

Fig. 2 | Waddington landscapes have been used to visualize the dynamics of cell differentiation, with the differentiated states as the minima of the landscape; they constitute a low-dimensional representation of cell fate decisions. The figures show two distinct potential landscapes that could putatively describe the differentiation of a cell into two alternative types. These constitute different geometrical possibilities for the low-dimensional behavior. **A** There is a direct path from the central attractor to the two alternative attractors. **B** On exiting the central attractor, the flow goes toward the second saddle point and is subsequently diverted toward one of the attractors. Adapted from ref. 68.



William Gilpin. Modern machine learning models exhibit the surprising ability to perform unseen logical tasks, like mathematical proofs, despite having never encountered them during training. This reasoning capability only emerges once models reach a sufficient number of parameters, implying that scale alone enables generalization capability. In such settings, dimensionality is an asset, not a curse. Why wouldn't the same be true for biology? The brain distributes low-dimensional computations across thousands of neurons, while developmental pathways encode differentiation across hundreds of interacting genes. As in statistical learning, do these high-dimensional representations confer robustness, allowing biological systems to easily adapt to new tasks? Emerging ideas from mechanistic interpretability in large-scale foundation models, like induction heads or attention routing, may thus have informative analogs in high-dimensional biological systems.

Archishman Raju. Organismal Development is shaped by processes that involve complex gene regulatory networks and signaling pathways, which have been characterized in great detail in model organisms over the past several decades. However, our ability to quantitatively understand high-dimensional gene expression data remains limited, and we continue to rely on heuristic computational techniques to understand this data. This difficulty was, in many ways, anticipated by the work of a section of early embryologists and theoretical biologists who established a connection between development and dynamical systems theory and were pessimistic about the usefulness of a reductionist approach to developmental phenomena. Dynamical systems theory has a long tradition of first qualitatively characterizing the behavior of a system of differential equations and then quantitatively understanding the simplest “normal form” equations consistent with the qualitative behavior, which may characterize the universal behavior of a multiplicity of biological mechanisms. These mathematical insights need to be converted into calculational tools that can be directly connected to biological data. Recent attempts to formalize Waddington landscapes (Fig. 2) in the context of cell fate specification fall into this broader project^{17,45,67,68}. It is possible that these ideas are applicable in other contexts, particularly in cases where an underlying high-dimensional system has low-dimensional outcomes.

Shaon Chakrabarti. What constitutes ‘good’ low-dimensional representations of complex datasets and how to achieve such latent space embeddings, are interesting open questions⁶⁹. Recent advances in generative AI, such as Variational Autoencoders (VAEs)⁷⁰ and Generative Adversarial Networks (GANs)⁷¹, provide interesting solutions to these questions in the context of identifying biological cell-states, by disentangling multiple sources of variation in low-dimensional spaces^{72,73}.

However, it remains unclear how functionally relevant these inferred cell-states are, especially when the ground truth is unknown (as is typically the case). It also remains unclear how biological and measurement noise affect the cellular resolution achievable in the inference of these cell-state embeddings. While cell-states emerge from many biological processes working together, current measurement technologies can only access a small fraction of these processes simultaneously. Random fluctuations in these few measured modalities may therefore blur the boundaries between cell-states and set fundamental limits to their identification.

For example, such a limitation has recently been demonstrated in the specific context of the mammalian circadian clock, where cells in different clock phases correspond to biologically relevant (ground truth) cell-states⁷⁴. While population-averaged RNA levels can be used to learn these correct cell-states on a circular 2D latent space, single-cell RNA measurements are too noisy and generate incorrect latent spaces. This is true even if technical noise associated with the measurement technique is minimized⁷⁴. This, therefore, serves as an example of how biological noise can affect the information content of low-dimensional embeddings generated by deep learning algorithms. Here, approaches from Information Theory could provide deeper insight into the performance of these algorithms, by quantifying how much of the information content of the high-dimensional cell-states is captured in the latent space. Information Theory could also be leveraged to predict how much further increase in the latent space information content might be theoretically possible, which could guide improvements in the deep neural network architecture and loss functions.

Ecology, evolution, and immunology

Michael Desai. Evolution is a process by which biology explores high-dimensional genotype-phenotype landscapes through random mutations, stochastic genetic drift, and natural selection. It might seem impossible that this process could reliably lead to adaptation in the face of diverse and ever-changing biological challenges. Yet evolution can solve many of these problems routinely and repeatably. How and why does this work? It may be that, instead of representing an obstacle, the high dimensionality of genotype and phenotype spaces leads to predictable emergent regularities that evolution can reliably exploit. Further, evolution may have led to regions of these spaces that have properties that are amenable to further adaptation. There may be precise connections that one can draw between these questions and the training of large machine learning models, and there is the potential for practical lessons from evolution to influence machine learning and vice versa. In particular, questions about the emergence and consequences of various forms of modularity were highlighted in the workshop and are clearly broadly important across these fields.

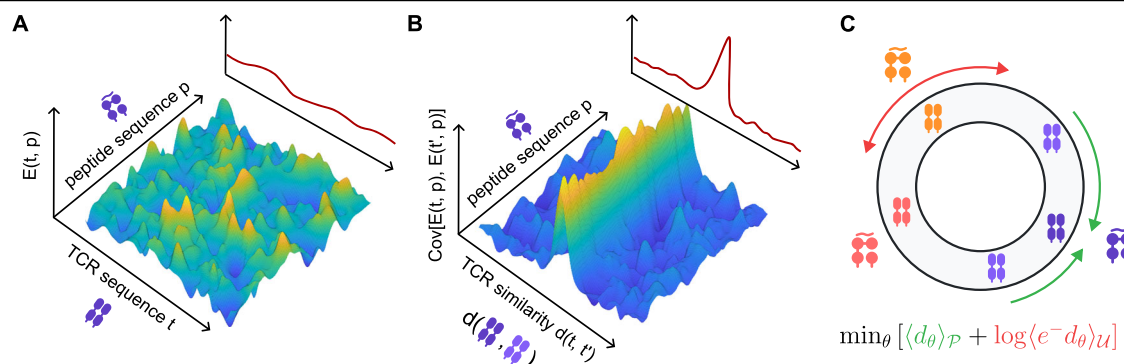


Fig. 3 | Contrastive learning can reveal generalizable similarity rules in the complex sequence to function map of immune receptors and their ligands.

A Cartoon of a binding energy landscape $E(t, p)$ describing T cell receptor t specificity to peptide ligands p . **B** Cartoon of the covariance of this energy between pairs of receptors t and t' as a function of the sequence similarity $d(t, t')$ between these pairs. By ‘aligning’ the peaks in the landscape, the co-variance structure is expected to generalize to a greater extent across ligands, as illustrated by the marginal

distributions (insets in **A** and **B**). **C** Contrastive learning optimizes parameters θ of a family of receptor similarity metrics $d_\theta(t, t')$ to minimize the distance between pairs of receptors \mathcal{P} that are co-specific to a common ligand, while encouraging uniform use of the representation space through a partition-function-like term involving an average over the null distribution of possible receptor pairs \mathcal{U} created by genetic recombination. [Adapted from ref. 21].

Jacopo Grilli. High-dimensionality is both the central challenge and the defining feature of community ecology—how do hundreds of species coexist despite complex interactions? A crucial distinction is that, unlike neural or cellular systems where optimality principles (minimizing energy, maximizing information processing) provide powerful organizing frameworks, ecological communities lack any obvious optimization target—there is no fitness function for a community, only for individuals within it. This raises fundamental questions: under what conditions can individuals actually behave optimally in high-dimensional environments where the fitness landscape constantly shifts due to other species? What even defines optimality is context-dependent (growth rate, yield, robustness), and the mismatch between individual-level optimization and community-level patterns is itself a source of emergent complexity. The challenge is compounded by dynamical feedback between organisms and environment, where both are high-dimensional: organisms possess many traits, and the environment comprises many factors, creating a complex co-evolutionary landscape where causes and effects are deeply entangled. Techniques from statistical physics and random matrix theory—scaling arguments, disorder averaging, studying typical rather than specific cases—offer ways forward by extracting universal patterns (macroecological laws, stability conditions) that emerge precisely because high-dimensionality washes out microscopic details. Conversely, ecology exports this crucial insight to other high-dimensional systems: the interplay between individual optimization, environmental feedback, and emergent collective behavior may be more fundamental than seeking system-level optimality principles, with applications from immune repertoires to artificial ecosystems where no designer’s objective exists.

Akshit Goyal. There is a longstanding tradition in physics of replacing complexity with randomness when it comes to complex, high-dimensional systems. This approach traces back to Wigner and Dyson’s seminal work in the 50s, who used random matrix theory to successfully predict statistical properties of the uranium nucleus^{75,76}. The key insight has always been to use randomness while respecting the fundamental constraints of the system of interest; for Wigner, this meant preserving the symmetry of the random Hamiltonian. May adopted this philosophy remarkably early in ecology (1972), modeling interspecies interactions as random matrices to

probe the stability of complex, species-rich ecosystems⁷⁷. Because ecological dynamics operate with a large number of species, the random matrix approach offered a tractable route to navigate high-dimensional complexity. Over the past decade, this idea has had a resurgence, resulting in new predictions about many statistical properties of ecosystems^{78–80}.

Yet a persistent criticism remains: these approaches use the “wrong” random ansatz, with the familiar refrain from biologists that “ecosystems are not random”. While everyone agrees that species and their interactions possess structure, determining what this structure should generically look like without imposing it by hand has been challenging. This concern extends across biological systems, e.g., intracellular condensates, metabolic networks, and other contexts where randomness has been invoked.

An emerging resolution is that for biological systems, the appropriate constraints may arise from evolution itself. In preliminary work on eco-evolutionary dynamics, we find that even when communities begin with random, unstructured interactions, continual feedback between ecological and evolutionary processes produces communities that remain diverse yet develop highly structured interaction patterns⁸¹. Surviving strains become strongly phenotypically correlated, and the evolved interaction structure is consistent and predictable. This work gives us hope that evolutionary stability is a missing ingredient that we must account for when thinking about ecosystems. The implication is potentially unifying: disordered interactions with evolutionarily stable structure may provide the “right” ansatz for studying high-dimensional biological systems.

Andreas Tiffeau-Mayer. Immunology is a paragon of high-dimensional complexity in biology. The trillion lymphocytes in our bodies carry more than a hundred million different antigen receptors drawn at random from an even more diverse set of possibilities. Yet, luckily, most of us are typically able to mount an immune response to any novel pathogen challenge. This crucially hinges on the multitude of receptors that bind any given target. Recent progress in deciphering the many-to-many mapping between immune receptors and pathogens has begun to reveal the biophysical rules underlying this redundancy in the coverage of antigen space.

As in other areas of protein biophysics, contrastive learning has emerged as a key strategy for learning function-aware coarse-graining procedures (Fig. 3). In the immunological context, this approach has

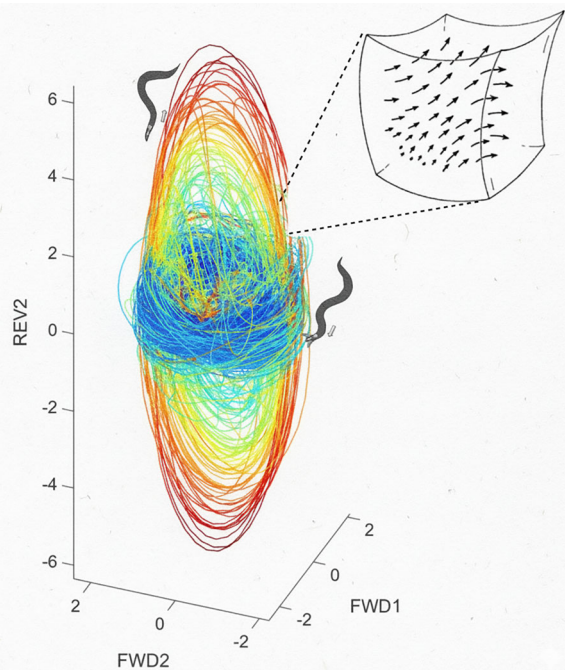


Fig. 4 | Low-dimensional dynamics in the behavior of *C. elegans*. We show a 3D projection of embedded worm posture trajectories illustrating forward (blue) and backward (red) locomotion through approximately orthogonal collections of cycles, a simplification of the full 7D state space⁴¹. Cartoon arrows are a reminder that the geometry of this space carries fundamental dynamical information, such as the Lyapunov exponents. Figure credit: Tosif Ahamed. [For details on the full 7D state space, see ref. 41].

uncovered statistical rules about which receptors are likely to bind the same ligand, revealing, for example, the importance of steric constraints in these interactions²¹. Importantly, knowledge of such rules enables detecting responses in immune repertoires by identifying convergently selected clusters of receptors with related binding specificity. Taming the immense receptor diversity by such coarse-graining approaches has thus emerged as a key advance in the study of the adaptive immune system, with applications in precision immunomonitoring and vaccine design.

Yet, if the aim is to develop a global theory of immune repertoire ecology, the description of immunological diversity at the level of these clusters of cells remains too high-dimensional for the identifiability of parameters from current experimental data. Therefore, techniques replacing detailed assumptions about interactions between immune cells and pathogens with random ensembles of fluctuating forces also play an important role in immunology^{82,83}. A potential future direction lies in the application of techniques developed for the analysis of longitudinal microbiome dynamics to reveal 'guild structure', pinpointing co-regulation of immune cells at scales beyond shared antigen specificity.

Behavior, learning, and larger scales

Greg Stephens. In the physical study of animal behavior, dimensionality has repeatedly emerged as a key organizing principle. Across diverse organisms, the space of body postures sampled from freely-moving conditions is low-dimensional, sometimes dramatically so, especially when compared to the number of underlying components, such as neurons or

muscles, from which behavior emerges. A classic example is the nematode worm *C. elegans*, where four "eigenworm" directions dominate the shape covariance matrix⁸⁴, even for a wide variety of mutants⁸⁵.

However, a more contentious and contemporary issue is the dimensionality of behavioral dynamics—how body configurations change over time. For example, recent analyses of *C. elegans* foraging have shown that worm behavior can be effectively described as a low-dimensional (seven!) chaotic dynamical system⁴¹ (see Fig. 4). Further, the complexity of this behavioral attractor can be manipulated; for example, six additional dimensions emerge from behavior in environments seeded with *E. coli*, but these can then be eliminated with an appropriate mutation.

In this view, complex posture dynamics and multiple timescales arise from low-dimensional but nonlinear interactions, which are reminiscent of the onset of turbulence⁸⁶. In contrast, long-term recordings of walking flies reveal an apparent power-law dynamics⁸⁷. Such power-law dynamics are at odds with a Markovian or low-dimensional framework, and one proposed resolution is that they result from time-varying internal states⁸⁸.

Another intriguing possibility, however, is that these findings in behavior mimic our early (mis)understanding of critical phenomena. For example, in the 2D Ising model, the mean-field magnetization undergoes an order-disorder transition, but with the wrong critical exponents - this is fixed only by considering the presence of (high-dimensional) fluctuations. Interestingly, this dimensional dichotomy is mirrored in neuroscience: while some studies point to low-dimensional neural manifolds (see, e.g., ref. 89), others reveal power-law activity⁹⁰. Resolving the tension between low-dimensional attractors and power-law behavior has the potential to reveal fundamental principles of both behavior and neural computation (see, e.g., ref. 91).

Simone Pigolotti. Throughout its history, physics has achieved remarkable success in describing a wide range of phenomena with elegant, simple, and accurate mathematical models. Notable examples include Newton's laws, Maxwell's equations, and the standard model of particle physics. Despite these successes, biology presents a formidable challenge. Even the simplest cells contain thousands of genes with different functions. This daunting complexity makes it unclear whether theory in biology will ever be as successful as in physics.

Two recent developments have dramatically changed the game. First, modern experimental techniques provide access to an unprecedented wealth of high-dimensional, quantitative data on biological systems. Second, the AI revolution has granted us tools to extract patterns from these data, and reduce these high-dimensional datasets to relatively few relevant variables.

These innovations offer a rather direct approach to our problem: gather as much high-quality data as possible and let AI discover the "model" for us. However, the reduced descriptions found by AI are not always easy to interpret. Moreover, not all modern biological data are high-throughput and accurate: in many areas, we are still dealing with limited and noisy datasets.

Achieving a deep understanding of biology will likely require a clever combination of traditional theory and data-driven AI methods. How to best integrate these two approaches remains an exciting open question.

Gautam Reddy. Much of the complexity of behavior arises from the fact that an animal's decision to make a move or take a path is fundamentally intertwined with the statistics of the environment it interacts with. Survival requires informed decisions, and informed decisions require an animal to store a representation of its history of movements and observations. But movements and observations are often "high-dimensional," and the number of possible histories is exceedingly large. Fortunately, most

information is not useful for behavior, and it is reasonable to believe that animals form representations that are useful specifically for guiding future, rewarding behavior rather than reconstructing past histories.

Representations of an animal's history and its own drives (such as hunger) form a persistent internal state. What is the “dimensionality” of this internal state space? How are internal states represented in neural systems? And how do they modulate context-dependent behavior? The capacity to precisely measure behavior, neural activity, connectomics, single-cell transcriptional states, and other modalities of neural communication offers hope that some of these questions will be resolved in the near future.

In parallel, recent developments in large language models (LLMs) offer an intriguing alternative path. LLMs do indeed develop causally identifiable internal states, and this capacity “emerges” after training large models on large amounts of data to predict what comes next, that is, without explicit human intervention. The computational substrate on which LLMs operate is, of course, different from biology, and the mechanistic details of how an internal state may arise will certainly be different. But it is natural to ask whether one can draw broader conceptual insights into the nature of internal states from a rigorous empirical study of LLM behavior.

Conclusion

As the datasets that describe the biological systems increase both in quality and dimensionality, several theoretical approaches toward understanding seem to emerge. One set of approaches targets the identification of an effectively low-dimensional description, which could, for some systems, justify and lead to the development of simple predictive microscopic models. Practically, a low-dimensional description can be found by compression of the larger space, or the identification of a representative latent space. This low-dimensional structure could also be described with a sloppy or soft mode framework. An interesting question is how such effective low-dimensionality is present across scales. Hypotheses that may explain this include stereotyped evolutionary or developmental trajectories that only allow the system to explore a limited part of state space. Alternatively, internal feedback or external constraints and performance optimizations could have driven the system toward specific subspaces.

Another set of approaches toward a unifying theory lies in embracing the high dimensionality: for example, by exploiting access to the statistics to fit a description or train a model. This high-dimensionality can apply both to a particular state and to the dynamics. Dynamic trajectories can retain high-dimensionality and partially chaotic behavior, which can nevertheless be quantified; alternatively, they may remain high-dimensional because they track an externally changing landscape. In this context, the dimensionality can be an asset to both biological and computational systems that allows them to fulfill multiple tasks using a given set of variables.

Future directions of research could involve an improved quantification of the effective dimensionality in the system and its dynamics, identifying drivers that lead to the emergence of specific structured sub-spaces in a specific biological question, and develop analogies between biological and artificial systems across scales.

Data availability

No datasets were generated or analyzed during the current study.

Marianne Bauer¹ ✉, **Akshit Goyal**², **Sidhartha Goyal**³, **Gautam Reddy**⁴, **Shaon Chakrabarti**⁵, **Michael M. Desai**⁶, **William Gilpin**⁷, **Jacopo Grilli**⁸, **Kabir Husain**⁹, **Sanjay Jain**^{10,11}, **Mohit Kumar Jolly**¹², **Kyogo Kawaguchi**^{13,14}, **Aneta Koseska**¹⁵, **Milo Lin**¹⁶, **Leelavati Narlikar**¹⁷, **Simone Pigolotti**¹⁸, **Archishman Raju**⁵, **Krishna Shrinivas**¹⁹,

Rahul Siddharthan^{20,21}, **Greg J. Stephens**^{22,23}, **Andreas Tiffeau-Mayer**²⁴ & **Suriyanarayanan Vaikuntanathan**^{25,26}

¹Bionanoscience Department, Kavli Institute of Nanoscience, Delft University of Technology, Delft, the Netherlands. ²International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru, India. ³Department of Physics, University of Toronto, Toronto, ON, Canada. ⁴Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ, USA. ⁵National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, India. ⁶Departments of Organismic and Evolutionary Biology and of Physics, Harvard University, Cambridge, MA, USA. ⁷Department of Physics, University of Texas at Austin, Austin, TX, USA. ⁸The Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy. ⁹Department of Physics and Astronomy, and Laboratory for Molecular Cell Biology, University College London, London, UK. ¹⁰Department of Physics and Astrophysics, University of Delhi, Delhi, India. ¹¹Santa Fe Institute, Santa Fe, NM, USA. ¹²Department of Bioengineering, Indian Institute of Science, Bengaluru, India. ¹³Nonequilibrium Physics of Living Matter Laboratory, Pioneer Research Institute, RIKEN, Kobe, Japan. ¹⁴Institute for Physics of Intelligence, Department of Physics, The University of Tokyo, Tokyo, Japan. ¹⁵Cellular Computations and Learning, Max Planck Institute for Neurobiology of Behavior-caesar, Bonn, Germany. ¹⁶Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA. ¹⁷Departments of Data Science and Biology, Indian Institute of Science Education and Research, Pune, India. ¹⁸Biological Complexity Unit, Okinawa Institute of Science and Technology, Onna, Okinawa, Japan. ¹⁹Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA. ²⁰The Institute of Mathematical Sciences, Chennai, India. ²¹Homi Bhabha National Institute, Mumbai, India. ²²Biological Physics Theory Unit, OIST Graduate University, Okinawa, Japan. ²³Department of Physics and Astronomy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ²⁴Division of Infection and Immunity & Institute for the Physics of Living Systems, University College London, London, UK. ²⁵The James Franck Institute, The University of Chicago, Chicago, IL, USA. ²⁶Department of Chemistry, The University of Chicago, Chicago, IL, USA.

✉ e-mail: m.s.bauer@tudelft.nl

Received: 11 December 2025; Accepted: 23 February 2026;

Published online: 28 March 2026

References

- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289 (2009).
- Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377 (2009).
- Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371 (2014).
- Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743 (2012).
- Sureshchandra, S. et al. Deep profiling of human T cells defines compartmentalized clones and phenotypic trajectories across blood and tonsils. *Immunity* **58**, 3130–3143.e8 (2025).
- Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457 (2017).
- Sunagawa, S. et al. Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* **18**, 428 (2020).
- The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207 (2012).
- Angelaki, D. et al. A brain-wide map of neural activity during complex behaviour. *Nature* **645**, 177 (2025).
- McKenzie-Smith, G. C., Wolf, S. W., Ayroles, J. F. & Shaevitz, J. W. Capturing continuous, long timescale behavioral changes in *Drosophila melanogaster* postural data. *PLoS Comput. Biol.* **21**, e1012753 (2025).

11. Gutenkunst, R. N. et al. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189 (2007).
12. Transtrum, M. K., Machta, B. B. & Sethna, J. P. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E* **83**, 036701 (2011).
13. Transtrum, M. K. & Qiu, P. Model reduction by manifold boundaries. *Phys. Rev. Lett.* **113**, 098701 (2014).
14. Scott, M. et al. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099 (2010).
15. Scott, M. & Hwa, T. Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nat. Rev. Microbiol.* **21**, 327 (2023).
16. Corson, F. & Siggia, E. D. Gene-free methodology for cell fate dynamics during development. *eLife* **6**, e30743 (2017).
17. Rand, D. A., Raju, A., Sáez, M., Corson, F. & Siggia, E. D. Geometry of gene regulatory dynamics. *Proc. Natl. Acad. Sci. USA* **118**, e2109729118 (2021).
18. Petti, S., Reddy, G. & Desai, M. M. Inferring sparse structure in genotype–phenotype maps. *Genetics* **225**, iyad127 (2023).
19. Johnson, M. S., Reddy, G. & Desai, M. M. Epistasis and evolution: recent advances and an outlook for prediction. *BMC Biol.* **21**, 120 (2023).
20. Van Den Brink, S. C. et al. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* **582**, 405 (2020).
21. Pyo, A. G. et al. Data-driven discovery of biophysical T cell receptor cospecificity rules. *PRX Life* **3**, 033005 (2025).
22. Meshulam, L. & Bialek, W. Statistical mechanics for networks of real neurons. *Rev. Mod. Phys.* **97**, 045002 (2025).
23. Zhong, L. et al. Unsupervised pretraining in biological neural networks. *Nature* **644**, 741–748 (2025).
24. Floyd, C., Dinner, A. R., Murugan, A. & Vaikuntanathan, S. Limits on the computational expressivity of non-equilibrium biophysical processes. *Nat. Commun.* **16**, 7184 (2025).
25. Chari, T. & Pachter, L. The specious art of single-cell genomics. *PLoS Comput. Biol.* **19**, e1011288 (2023).
26. Grobceker, P., Sakoparnig, T. & van Nimwegen, E. Identifying cell states in single-cell RNA-seq data at statistically maximal resolution. *PLoS Comput. Biol.* **20**, e1012224 (2024).
27. Furlong, E. E. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341 (2018).
28. Bintu, L. et al. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15**, 125 (2005).
29. Scholes, C., DePace, A. H. & Sánchez, Á. Combinatorial gene regulation through kinetic control of the transcription cycle. *Cell Syst.* **4**, 97 (2017).
30. Tkačik, G., Callan Jr, C. G. & Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA* **105**, 12265 (2008).
31. Bauer, M., Petkova, M., Gregor, T., Wieschaus, E. F. & Bialek, W. Trading bits in the readout from a genetic network. *Proc. Natl. Acad. Sci. USA* **118**, e2109011118 (2021).
32. Tkačik, G. & ten Wolde, P. R. Information processing in biochemical networks. *Annu. Rev. Biophys.* **54**, 249 (2025).
33. Sokolowski, T. R., Gregor, T., Bialek, W. & Tkačik, G. Deriving a genetic regulatory network from an optimization principle. *Proc. Natl. Acad. Sci. USA* **122**, e2402925121 (2025).
34. Mijatović, T. et al. Weak transcription factor clustering at binding sites can facilitate information transfer from molecular signals. *PRX Life* **4**, 013003 (2026).
35. Bialek, W. *Biophysics: Searching for Principles* (Princeton University Press, 2012).
36. Bialek, W., van Steveninck, R. R. D. R. & Tishby, N. Efficient coding in sensory systems. In *Proc. IEEE International Symposium on Information Theory* 659–663 (IEEE, 2016).
37. Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* **36**, 910 (1981).
38. Bauer, M. et al. Optimization and variability can coexist. Preprint at <https://arXiv.org/abs/2505.23398> (2025).
39. Witteveen, O. et al. Optimizing information transmission in the canonical Wnt pathway. *Phys. Rev. Res.* **8**, 01329 (2026).
40. Selimkhanov, J. et al. Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370 (2014).
41. Ahamed, T., Costa, A. C. & Stephens, G. J. Capturing the continuous complexity of behaviour in *Caenorhabditis elegans*. *Nat. Phys.* **17**, 275 (2021).
42. Moor, A.-L. & Zechner, C. Dynamic information transfer in stochastic biochemical networks. *Phys. Rev. Res.* **5**, 013032 (2023).
43. Reinhardt, M., Tkačik, G. & ten Wolde, P. R. Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories. *Phys. Rev. X* **13**, 041017 (2023).
44. Reinhardt, M. et al. Mutual information rate–linear noise approximation and exact computation. Preprint at <https://arXiv.org/abs/2508.21220> (2025).
45. Raju, A. & Siggia, E. D. A geometrical perspective on development. *Dev. Growth Differ.* **65**, 245 (2023).
46. Gilpin, W. Generative learning for nonlinear dynamics. *Nat. Rev. Phys.* **6**, 194 (2024).
47. Yadav, M., Koch, D. & Koseska, A. Homeorhetic regulation of cellular phenotype. Preprint at <https://arXiv.org/abs/2506.06658> (2025).
48. Nguyen, A. & Reddy, G. Differential learning kinetics govern the transition from memorization to generalization during in-context learning. International Conference for Learning Representations. (2025).
49. Zentner, A. et al. Information processing driven by multicomponent surface condensates. Preprint at <https://arXiv.org/abs/2509.08100> (2025).
50. Rukhlenko, O. S. et al. Control of cell state transitions. *Nature* **609**, 975 (2022).
51. Zhou, P., Wang, S., Li, T. & Nie, Q. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. *Nat. Commun.* **12**, 5609 (2021).
52. Xing, J. & Kim, K. S. Application of the projection operator formalism to non-Hamiltonian dynamics. *J. Chem. Phys.* **134**, 044132 (2011).
53. Chen, Y. et al. GraphVelo allows for accurate inference of multimodal velocities and molecular mechanisms for single cells. *Nat. Commun.* **16**, 7831 (2025).
54. Wang, W., Ni, K., Poe, D. & Xing, J. Transiently increased coordination in gene regulation during cell phenotypic transitions. *PRX Life* **2**, 043009 (2024).
55. Hari, K. et al. Low dimensionality of phenotypic space as an emergent property of coordinated teams in biological regulatory networks. *iScience* **28**, 111730 (2025).
56. Qiu, X. et al. Mapping transcriptomic vector fields of single cells. *Cell* **185**, 690 (2022).
57. Islam, S. & Bhattacharya, S. Dynamical systems theory as an organizing principle for single-cell biology. *npj Syst. Biol. Appl.* **11**, 85 (2025).
58. Samal, A. & Jain, S. The regulatory network of *E. coli* as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC Syst. Biol.* **2**, 21 (2008).
59. Pandey, P. P., Singh, H. & Jain, S. Exponential trajectories, cell size fluctuations, and the adder property in bacteria follow from simple chemical dynamics and division control. *Phys. Rev. E* **101**, 062406 (2020).
60. Freedman, S. L., Xu, B., Goyal, S. & Mani, M. A dynamical systems treatment of transcriptomic trajectories in hematopoiesis. *Development* **150**, dev201280 (2023).
61. Farrell, S., Mani, M. & Goyal, S. Inferring single-cell transcriptomic dynamics with structured latent gene expression dynamics. *Cell Rep. Methods* **3**, 100479 (2023).
62. Shakiba, N. et al. Cell competition during reprogramming gives rise to dominant clones. *Science* **364**, eaan0925 (2019).
63. Ryu, H. et al. Frequency modulation of ERK activation dynamics rewires cell fate. *Mol. Syst. Biol.* **11**, 838 (2015).
64. Rosen, S. et al. Anti-resonance in developmental signaling regulates cell fate decisions. *eLife* **14**, RP107794 (2026).
65. Kloeden, P. E. & Rasmussen, M. *Nonautonomous Dynamical Systems* Vol. 176 (American Mathematical Society, 2011).
66. Nandan, A., Das, A., Lott, R. & Koseska, A. Cells use molecular working memory to navigate in changing chemoattractant fields. *eLife* **11**, e76825 (2022).
67. Sáez, M. et al. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Syst.* **13**, 12 (2022).
68. Raju, A. & Siggia, E. D. A geometrical model of cell fate specification in the mouse blastocyst. *Development* **151**, dev202467 (2024).
69. Bengio, Y., Courville, A. & Vincent, P. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798 (2013).
70. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. In *Proc. International Conference on Learning Representations (ICLR)*, 2014.
71. Goodfellow, I. J. et al. Generative adversarial networks. In *Proc. Advances in Neural Information Processing Systems* 2672–268 (NeurIPS, 2014).
72. Lopez, R. et al. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053 (2018).
73. Marouf, M. et al. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat. Commun.* **11**, 166 (2020).
74. Nikhat, A. et al. Transcriptional noise sets fundamental limits to decoding circadian clock phase from single-cell RNA snapshots. <https://doi.org/10.1016/j.jsci.2026.115394> (iScience, 2026).
75. Wigner, E. P. On the statistical distribution of the widths and spacings of nuclear resonance levels. *Math. Proc. Camb. Phil. Soc.* **47**, 790–798 (1951).
76. Dyson, F. J. Statistical theory of the energy levels of complex systems. *I. J. Math. Phys.* **3**, 140 (1962).
77. May, R. M. Will a large complex system be stable? *Nature* **238**, 413 (1972).
78. Nemenman, I. & Mehta, P. Randomness with constraints: constructing minimal models for high-dimensional biology. Preprint at <https://arXiv.org/abs/2509.03765> (2025).
79. Cui, W., Marsland III, R. & Mehta, P. Les Houches lectures on community ecology: From niche theory to statistical mechanics. Preprint at <https://arXiv.org/abs/2403.05497> (2024).
80. Bunin, G. Ecological communities with Lotka–Volterra dynamics. *Phys. Rev. E* **95**, 042414 (2017).
81. Feng, Z., Blumenthal, E., Mehta, P. & Goyal, A. A theory of ecological invasions and its implications for eco-evolutionary dynamics. *Proc. Natl. Acad. Sci. USA* **122**, e250585012 (2025).
82. Desponds, J., Mora, T. & Walczak, A. M. Fluctuating fitness shapes the clone-size distribution of immune repertoires. *Proc. Natl. Acad. Sci. USA* **113**, 274 (2016).
83. Gaimann, M. U., Nguyen, M., Desponds, J. & Mayer, A. Early life imprints the hierarchy of T cell clone sizes. *eLife* **9**, e61639 (2020).
84. Stephens, G. J., Johnson-Kerner, B., Bialek, W. & Ryu, W. S. Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput. Biol.* **4**, e1000028 (2008).
85. Brown, A. E. X. et al. A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *Proc. Natl. Acad. Sci. USA* **110**, 791 (2013).
86. Brandstätter, A. et al. Low-dimensional chaos in a hydrodynamic system. *Phys. Rev. Lett.* **51**, 1442 (1983).
87. Bialek, W. & Shaevitz, J. W. Long timescales, individual differences, and scale invariance in animal behavior. *Phys. Rev. Lett.* **132**, 048401 (2024).
88. Costa, A. C., Sridhar, G., Wyart, C. & Vergassola, M. Fluctuating landscapes and heavy tails in animal behavior. *PRX Life* **2**, 023001 (2024).

89. Perich, M. G., Narain, D. & Gallego, J. A. A neural manifold view of the brain. *Nat. Neurosci.* **28**, 1582–1597 (2025).
90. Manley, J. et al. Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron* **112**, 1694 (2024).
91. Fontenele, A. J. et al. Low-dimensional criticality embedded in high-dimensional awake brain dynamics. *Sci. Adv.* **10**, ead9303 (2024).

Acknowledgements

This research was supported in part by the International Centre for Theoretical Sciences (ICTS) for participating in the program - Unifying Theories in High-Dimensional Biophysics (code: ICTS/UHDB2025/07), as well as the American Physical Society DBIO (Division of Biological Physics). This work was also supported by Fondo Italiano per la Scienza - FIS (CUP J53C23002290001) (JG), by JSPS KAKENHI Grant No. 23H01146 (SP) and by the Department of Atomic Energy, Government of India (under project RTI4006, AR).

Author contributions

M.B., A.G., S.G., G.R., S.C., M.M.D., W.G., J.G., K.H., S.J., M.K.J., K.K., A.K., M.L., L.N., S.P., A.R., K.S., R.S., G.J.S., A.T.M., S.V. contributed to drafting, editing, and final approval, and wrote their respective paragraph of this manuscript.

Competing interests

M.B., A.G., S.G., G.R., S.C., M.M.D., W.G., J.G., K.H., S.J., K.K., A.K., M.L., L.N., S.P., A.R., K.S., R.S., G.J.S., A.T.M., and S.V. declare no competing financial or non-financial interests. M.K.J. serves as

Editor in Chief of this journal and had no role in the peer review or decision to publish this manuscript. M.K.J. declares no competing financial interests.

Additional information

Correspondence and requests for materials should be addressed to Marianne Bauer.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026