

On the emergence of in silico biology for intervention in living systems

James Colter

jdcolter@ucalgary.ca

University of Calgary <https://orcid.org/0000-0001-6612-6931>

Nicole Rosin

nicole.rosin@ucalgary.ca

University of Calgary <https://orcid.org/0000-0003-3030-1704>

Jeff Biernaskie

University of Calgary <https://orcid.org/0000-0002-0633-0698>

Nils Daniel Forkert

University of Calgary <https://orcid.org/0000-0003-2556-3224>

Short Report

Keywords: Machine Learning, Systems Biology, In Silico Biology, Deep Learning

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11321044/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Advances in scale and depth across omics, biomarker, and imaging technologies now resolve molecular, cellular, tissue, and system-level biology in unprecedented detail. These data create vast opportunities to represent biological state, predict biological response, and guide experimental intervention in medicine. However, contemporary data science and human capacity for interpretation fall short at this scale.

Artificial intelligence has become a principal focus in attempting to overcome challenges in biological modeling. Frontier modelling approaches learn representations from heterogeneous, multimodal datasets, providing tools to organize, characterize, and attempt to predict biological states under defined contexts. However, the value of these models is thus far limited, with little evidence of translation. Progress beyond the contemporary depends on appropriate design of computation around the architecture of biology: distinguishing context-specific associations from relationships that remain predictive across biological conditions, encoding understanding of where observed correlations converge or diverge within mechanisms, and consolidating interactions across molecule, cell, tissue, and system.

In this work, we examine *in silico* biology from a systems perspective: computational systems that learn representations of biological state, predict state transitions, prioritize experimentally testable hypotheses, and guide intervention in living systems. We organize the field around three intertwined problems. The first is the representation problem: Learning biological state from heterogeneous, sparse, disconnected measurements. The second is the causality problem: Predicting how biological states maintain and respond to stimuli. The third is the intervention problem: Validating computationally perceived phenomena.

In our outlook, the next frontier is not defined by architecture, modality, or information scaleup alone. It will emerge through cohesive representation across contexts, domains, and scales; interrogation of causality against computational design within and beyond the data, and through iterative computational-experimental intervention to refine both data and architecture to fill the gaps that contextualize causality against biology broadly.

Introduction

Biology is now measured at unprecedented scale and resolution. Single-cell and spatial omics, multiplex imaging, perturbation screens, longitudinal clinical profiling, and population-scale datasets have transformed our capacity for discovery [1-6]. Two central challenges now pervade the field. The first, is in information synthesis, and the second is in resolving cause and effect. Computational biology has witnessed a profound shift towards architecting artificial systems to overcome these challenges, across a wide range of philosophical approaches [7-9].

Context is critical; many consequential biological and medical questions are intervention problems. Causal hierarchy distinguishes association, which asks what is observed; intervention, which asks what

will happen if a variable is changed; and counterfactual reasoning, which ponders response under alternative action [10]. Biological datasets have predominantly documented associations, but therapeutic decisions require prediction about intervention and comparison among plausible alternatives. A therapeutic target is valuable only if modulating it produces a desired change from a defined initial state. A biomarker is useful insofar as it informs diagnosis, prognosis, treatment selection, or response. A regenerative therapy must induce a healthy state in the target niche while supporting an appropriate return to and maintenance of healthy homeostasis. A disease model must identify reversible cascades, mechanisms of priming and activation, and compensatory responses that limit intervention to facilitate a reversal in phenotype. *In silico* biology must climb the causal hierarchy: from recognizing patterns in observation, to predicting the consequences of intervention, to reasoning how alternative action belies compensation, adaptation, and resolution.

The aforementioned problems center around context, but these problems share commonalities across generalizable phenomena within how biology regulates itself. However, the tasks of solving a contextual problem or generalizing mechanism across biological contexts are not so simple. Biology is governed by strict rulesets, across hundreds of thousands to millions of evolutionarily coordinated interactions to drive context, and outside of that context contribute to a generalizable network. Computationally then, context simplifies *in silico* architectures to great benefit for the accuracy of a model, yet may have limited applicability across the breadth of biology [11]. The question then is how do we reconcile the data that exists, the data we need, and the computational architectures designed around this space to facilitate models of biology whose predictions stand across the breadth of molecular tasks, cell types, tissues, systems, and environments that exist, let alone within a particular context?

We refer to this emerging paradigm of capturing cross-mechanistic biological structure and dynamics as *in silico* biology: computational systems that understand the causality of life foundationally across biological representation [12].

In line with understanding contextual dependence, the field's focus on experimental perturbation provides some measure of experimental currency, directly testing whether a model can predict how a constrained view of biology changes when specific variables are altered [13]. But what of the variables influencing organization of a phenotype under experimental interrogation? The epigenetic state under which entities are perturbed? The contribution of environmental compensation and adaptation to alternative mechanism? The lack of natural constraints normally applied by tissue or system? It is critical to consider the dichotomy of perturbation in conventional practice against reality [14, 15].

This Perspective argues that the principal barrier to *in silico* biology is neither deficiency in data, architecture, nor validation; rather the separation of representation, causality, and intervention into independent context, domain, and scale of complexity. A model cannot predict a transition governed by biology absent its encoded representation, and an experiment cannot establish causal relevance when the contexts driving naturally organized response, compensation, and adaptation are excluded from its design.

Herein we examine these interconnected problems. Firstly, the representation problem: learning biological state from heterogeneous, disconnected, partial multi-modal measurements. Second, the causality problem: predicting how biological states are maintained and altered under perturbation across regulatory domains and layers of organizational complexity. Third, the intervention problem: validating perceived causality when challenged in experimental, physiological, and clinical contexts.

I. The representation problem: Reconstructing biological state dynamics

Biology is measured through partial observations: epigenetic context; genomic sequence; downstream abundance and regulation; spatial organization and morphology; tissue organization; physiological readouts; clinical outcome. In humans, these measurements are fragmented across genomic screens, biomarker panels, patient samples, diagnostic data, clinical trajectories, treatment exposure, coexisting disorders, longitudinal response, and post-mortem analysis, all in the face of diverse lived experience and bias, systemic or otherwise [16] (Figure 1).

Each modality we measure captures a projection of biology upon a particular domain. The core question underlying the representation problem is how to inform the cohesion of programs that maintain and transition biology over contexts, domains, time-scales, conditions, and in response to stimuli from observational or perturbational snapshots. To what extent can our data inform causality within the scope of design? What of outside this scope? Can we infer beyond what is visible or hidden within existing data with any realistic relevance?

Conventional bioinformatics are broadly designed to integrate, summarize, highlight, and dissociate measured states [17]. Extensions including trajectory inference, dynamical modelling, network reconstruction, and causal inference estimate how states arise, progress, and respond to changes in feature inputs [18-20]. However, statistical approaches reconstruct covariance from snapshots and assumptions, replacing missing information about fundamental rules of governance with statistical distributions that are limited in context. Causal discovery provides formal methods for recovering candidate causal structure from observational and experimental data, yet these structures remain conditional on the variables measured, the assumptions imposed, and the contexts sampled. In biological systems, latent regulation, incomplete temporal observation, feedback, compensation, and context-dependent mechanisms leave multiple causal models artificially compatible with the same data [10, 21].

Take the AP-1 transcription factor subunit c-Jun as a principal example. This is a gene required for fundamental processes within a cell, yet its transcription responds contextually in governing cellular response across more than two thousand predicted interactions, few causal, with most indirect by covariance. It's transcription is implicated in both de-differentiation-like activity following injury, and exit from pluripotency, only two dichotomous contexts of many [22, 23]. In itself it is generally fundamental to many functions, but adjacently important in orchestrating contextually-dependent cellular response. Its role under context across predicted interactions is unclear. The underlying mechanisms are

transcriptionally indistinguishable in the absence of informed perturbation, counterfactual, and out-of-domain context.

Mechanistic priors, including sparse molecular interaction maps, gene-regulatory networks, signalling pathways, spatial neighbourhoods, lineage relationships, and evolutionary conservation offer constraint to guide models where direct measurements are absent, priors are unknown, or variability needs be organized around core biological ruleset. The representation problem then sets the boundaries within which the causality problem is constrained. How do we design observation to distinguish context-appropriate causality from that which is more broadly transferable? How do we design models to capitalize on this observational design? When technical and biological variation are confounded by experimental design, preprocessing or model architectures optimized for surrogate objectives may obscure genuine mechanistic information. Alignment achieved under these objectives may establish correspondence between observed states, but not capture the molecular interactions, spatial dependencies, population structure, or regulatory constraints through which they are generated and maintained [24-26].

Appropriate early model designs may draw on neurosymbolic approaches to structure representations before causal interrogation and experimental validation. These models combine subsymbolic learning with semantic networks, mechanistic rules, or other structured priors [27]. Such structure can constrain candidate causal models and support subsequent interventional and counterfactual reasoning, but does not itself establish causal validity.

Prior knowledge provides valuable constraints, particularly where direct observations are sparse, but it should not be treated as universal ground truth. Interaction and pathway resources aggregate evidence across heterogeneous tissues, organisms, assays, and disease states, and few capture validated causality, where contextualization across databases is uneven. If imposed without calibration against the target system, these priors can propagate irrelevant information, ignore sparsely represented mechanisms, and bias predictions [15].

This raises an important question: do we possess sufficient biological information to construct generalized models of cellular behavior? The continued accumulation of ever-larger transcriptomic datasets may offer diminishing returns relative to deeper characterization of complementary regulatory layers, including the epigenome, proteome, and the regulatory machinery that encompass them [28-30]. So too must our models effectively encode this information. Is it enough to capture the breadth of distributions by scaling the quantity and depth of data? Or should architectures more thoughtfully organize biology with coherence to underlying causality, explicitly or otherwise?

Frontier II. The causality problem: In silico prediction of biological response

While the representation problem concerns describing biological state, the causality problem directly informs prediction of biological change. Given an initial cellular, tissue, or organismal state, can a model

accurately predict how that system will respond to a previously unobserved perturbation, environment, or context?

Many modern models can successfully interpolate within observed biological distributions, particularly when architectural design and data utilization are well matched to the task. This capability alone provides substantial value, as the scale and dimensionality of modern biological datasets exceeds the capacity of human cognition. Yet if our scientific ambition extends beyond extracting causality from observation, and the ultimate goal is to predict causal response that has not yet been measured, how do we get there?

New models pretrained on massive molecular and single-cell datasets continue to pursue transferable biological prediction through zero-shot, few-shot and fine-tuned applications [31-33]. Yet their limitations raise a fundamental question: why is model scale treated as the principal bottleneck? Are these models not constrained by the biological information available for learning, the objectives through which that information is encoded, or both (Table 1)? Current evidence suggests that models trained predominantly on observational transcriptomes learn statistical structure that can support downstream prediction, but not necessarily causality, nor are they yet proven to generalize reliably to unseen context [34-36].

Biological Domain		Dominant Scale(s)				
		Molecular	Cell	Tissue	Patient	Population
Modeling Coverage						
Genome	Substantial	■	■	■	■	■
Epigenome	Fragmentary	■	■	■	■	
Transcriptome	Substantial	■	■	■	■	
Proteome	Fragmentary	■	■	■	■	
Metabolome	Sparse	■	■	■	■	
Morphology and Imaging	Partial	■	■	■	■	
Physiology	Partial			■	■	■
Clinical	Partial				■	■

Table 1: A qualitative synthesis of data availability for computation across domains and scales.

Whether fine-tuning can recover regulatory context, temporal dependencies, biochemical constraints, tissue organization, or compensatory responses depends on whether the relevant information is present in the observations and retained in the pretrained representation [34-37]. Further, abundant measurements of sequence, transcription, morphology, and clinical association remain disconnected or weakly connected to the perturbational, longitudinal, and multiscale evidence needed to validate biological response [38]. The transition problem thus inherits the representation problem: models cannot

reliably generalize responses across contexts where governing causality is absent, disconnected, or confounded across training tasks [39].

Perturbation experiments provide a more direct route towards resolving this limitation by measuring biological response under specified interventions. Such data can be used both to train models on intervention-response relationships and to validate causal hypotheses generated from existing observational data. Relative to observational associations, perturbations more strongly constrain candidate mechanisms, but they do not establish mechanism unambiguously: observed responses may reflect on-target activity, off-target effects, pleiotropy, compensation or conditional adaptation [30, 40]. Responses observed *in vitro* must then be re-evaluated within tissue microenvironments, under systemic regulation, across heterogeneous populations and, where model organisms are used, across species. Extending perturbational measurements across developmental stages, tissues, and species is non-trivial in the pursuit of broadly transferable rules [30, 38, 40].

Perturbation studies targeting gene expression nevertheless interrogate only a subset of the regulatory processes that shape biological responses. Predictive models need integrate evidence across genomic, epigenomic, transcriptomic, proteomic and metabolic layers of regulatory information, while connecting cellular interaction, tissue organization and systemic physiology [28, 41]. This may require architectures matched to the structure of the biological problem. Graph neural networks can encode entities and their known or inferred relationships, including molecular interactions, regulatory networks and spatial neighbourhoods, whereas transformer-based models can learn context-dependent and potentially long-range dependencies across large, heterogeneous datasets [41-43]. Hybrid architectures may provide a more reflective route towards combining relational priors with learned representations, though their mechanistic value still remains dependent on the observations and objectives used to train them.

Frontier III. The intervention problem: Validating perceived biological understanding

Computational pattern recognition of biological information has little relevance without applied validation. Experiments must define context, challenge predicted mechanisms, and generate observations that expose where proposed candidates succeed or fail. The resulting evidence can refine model parameters, objectives and biological priors, and support further determination of which predictions generalize beyond the contexts represented in the training data.

Biological validation as a stagewise approach first determines if a model predicts a defined outcome under specified experimental conditions. A successful test supports that prediction within the evaluated context, but does not by itself establish mechanism or transferability. These require additional experiments designed to test whether the prediction persists under counterfactual and mechanistic hypotheses across scales, to tissue microenvironment, organismal physiology and across a population [30, 44, 45]. The appropriate validation strategy is also dependent on what information is missing. Some variables can be measured directly, approximated through related modalities or tested through targeted perturbation; others may be inaccessible, unobserved or non-identifiable within the available experimental system. Where the experimental space is large, models can prioritize interventions that

discriminate among competing hypotheses or evaluate contexts in which predictions are most uncertain [21, 46]. The resulting experiments can then evaluate those predictions while providing new evidence for our models. In the meantime, current approaches need be ever mindful of the context in which co-correlates are not valid.

Take the virtual cell movement as a case study; available benchmarks do not establish a superior architecture across evaluated single-cell modelling tasks, with metrics further blurred by the definition of the task [30, 36, 47-49]. When evaluated without task-specific fine-tuning, pretrained transformer-based foundation models do not outperform highly variable gene selection or variational autoencoder-based approaches in cell-type clustering, batch integration, nor prediction tasks [34-36, 39, 48]. Perturbation-response benchmarks likewise show that relative model performance depends on the dataset, prediction task, and evaluation metric [46, 48, 49]. Importantly, agreement with observed expression profiles and the ability to distinguish perturbation-specific responses are separate evaluation criteria, while metric calibration can affect whether improvements over baseline predictions are detected. This is demonstrated by the variance in outcome across recent benchmarks and model releases (Figure 3a) [46, 48, 49]. Further, recent model releases such as STATE demonstrate substantial gains under particular evaluation settings without establishing uniform improvement across contexts and endpoints (Figure 3b) [50]. Given the state of our models under specified contexts across comparative analyses against well-established transcriptional abundance tasks, how far does the needle need to move to broadly facilitate consistent multi-domain, multi-scale implementations that hold up to validation?

Critical in these arguments is the distinction between latent transition, relational statistic, and explicit feature prediction relative to differential remodeling of a biological system. AP-1 again provides a wonderful system to make this point. Clearly core within many programs, under inappropriate constraints a perturbation to this program will rank highly in transitioning a system both towards and away from an intended target [51, 52]. It is simple enough to detract conflicting outcomes from a rank list of optimized target predictions, but less so is capturing the nonlinearity of that response relative to the causality of the system. Key to understand is the drivers propagating those responses across experimental system and scale relative to the model; what informative feedback optimizes the architecture and forwards the computational implementation?

The next Frontier: *in silico* biology

Living systems coordinate molecular regulation spanning cellular architecture, spatial organization, tissue dynamics, and physiology, across myriad contexts and environmental challenge. The next frontier of computational systems must do far more than organize biological variation around distributions tuned against feature sets to solely describe high-variant features differentiating entities and systems. This ambition is constrained less by the volume of biological data than by its fragmentation. Molecular, cellular, spatial, physiological, clinical, and population measurements provide partial, uneven views of connected systems. Current models do not sufficiently link initial state, intervention, mechanism, response, compensation, sameness, nor outcome across contexts, nor do they explicitly learn to

accomplish this across scales. Models consequently generalize across contexts with governing mechanisms absent, confounded, or disconnected from the endpoints they predict. This will work in many contexts, and is insufficient in many more. Increasing data will improve distributions, but without thoughtful experimental design poorly reflects or may entirely miss unobserved biological architecture.

Computational architecture must reflect this limitation. Current evidence does not identify a universally optimal model class, nor does increasing architectural complexity reliably establish certainty under context nor generalization. The relevant standard is whether a model preserves the structure required for its intended prediction and an understanding of how that accuracy diverges when challenged by the biology to which it is applied and further validated. Language models, structured latent encoders, dynamical systems, relational graphs, neurosymbolic systems, or simpler statistical models all show exciting promise in computational biology, and will no doubt contribute under appropriate design [53-57]. Their inclusion should be justified by performance under meaningful biological challenge, not by in-distribution reconstruction, token utilization, performance solely across the strongest defined feature variants, or benchmark optimization across low-hanging fruit. The next frontier will unify representation against causality against intervention, within novel computational-experimental scaffolds, across domains and scale. In the meantime, existing frameworks and philosophies must more consciously address the boundaries governing the biology over which we seek *in silico* control, to further the extent to which that control over biology truly extends.

Methods

Benchmark synthesis and visualization. Published single-cell perturbation benchmark results were consolidated across studies and grouped into four evaluation domains: expression fit, perturbation discrimination, distributional fidelity, and gene recovery. Within each benchmark metric, model performance was converted to a direction-aware relative standing score using rank normalization, allowing comparison of models across heterogeneous evaluation frameworks while preserving dataset-, task-, and metric-specific context. Domain scores were calculated only when all required metrics for that domain were reported. Context-transfer and combination-prediction tasks were evaluated separately and visualized concurrently. For STATE, published context-transfer results were represented as reported relative improvements over comparator models, whereas drug-combination performance was represented using reported BLISS and ZIP synergy correlations.

Declarations

Data Availability

All scripts and consolidated evidence for the secondary analyses and visualization performed for figure 3 are available at https://github.com/JxColter/ColterJ_et al_mlxbio_perspective.

Acknowledgements

The authors gratefully acknowledge the Canadian Institutes of Health Research and Alberta Innovates for supporting this work.

Ethics: Not applicable.

Consent to participate: Not applicable.

Consent to publish: Not applicable.

Funding: This work was funded a Canadian Institutes of Health Research Project Grant and by an Alberta Innovates Enhancement Postdoctoral Fellowship.

References

1. Gulati, G.S., D'Silva, J.P., Liu, Y. et al. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nat Rev Mol Cell Biol* 26, 11–31 (2025). <https://doi.org/10.1038/s41580-024-00768-2>
2. Lázár, E., Lundeberg, J. Spatial architecture of development and disease. *Nat Rev Genet* 27, 118–136 (2026). <https://doi.org/10.1038/s41576-025-00892-5>
3. Stupichev, D., Mihecheva, N., Postovalova, E. et al. AI-driven multimodal algorithm predicts immunotherapy and targeted therapy outcomes in clear cell renal cell carcinoma. *Cell Rep. Med.* 6, 102299 (2025). <https://doi.org/10.1016/j.xcrm.2025.102299>
4. Zheng, Y., Sadée, C., Ozawa, M., Howitt, B. E. & Gevaert, O. Single-cell multimodal analysis reveals tumor microenvironment predictive of treatment response in non-small cell lung cancer. *Sci. Adv.* 11, eadu2151 (2025). <https://doi.org/10.1126/sciadv.adu2151>
5. Gao, Y., Ventura-Diaz, S., Wang, X. et al. An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer. *Nat Commun* 15, 9613 (2024). <https://doi.org/10.1038/s41467-024-53450-8>
6. Replogle, J. M., Saunders, R. A., Pogson, A. N. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* 185, 2559-2575.e28 (2022). <https://doi.org/10.1016/j.cell.2022.05.013>
7. Cui, H., Tejada-Lapuerta, A., Brbić, M. et al. Towards multimodal foundation models in molecular cell biology. *Nature* 640, 623–633 (2025). <https://doi.org/10.1038/s41586-025-08710-y>
8. Rood, J. E., Hupalowska, A. & Regev, A. et al. Toward a foundation model of causal cell and tissue biology with a Perturbation Cell and Tissue Atlas. *Cell* 187, 4520-4545 (2024). <https://doi.org/10.1016/j.cell.2024.07.033>
9. Baião, A. R., Cai, Z., Poulos, R. C., et al. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Brief. Bioinform.* 26, bbaf355 (2025). <https://doi.org/10.1093/bib/bbaf355>

10. Prosperi, M. et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat. Mach. Intell.* **2**, 369–375 (2020).
11. Nagai, M., Murphy, A.E., Rizzo, K. et al. Toward generalizable and interpretable AI in regulatory genomics. *Nat Genet* **58**, 1787–1802 (2026). <https://doi.org/10.1038/s41588-026-02670-3>
12. Rao, V.M., Zhang, S., Plosky, B.S. et al. Generalist biological artificial intelligence in modeling the language of life. *Nat Biotechnol* **44**, 918–933 (2026). <https://doi.org/10.1038/s41587-026-03064-w>
13. Southard, K. M. et al. Comprehensive transcription factor perturbations recapitulate fibroblast transcriptional states. *Nat. Genet.* **57**, 2323–2334 (2025). <https://doi.org/10.1038/s41588-025-02284-1>
14. Webster, A.K., Phillips, P.C. Epigenetics and individuality: from concepts to causality across timescales. *Nat Rev Genet* **26**, 406–423 (2025). <https://doi.org/10.1038/s41576-024-00804-z>
15. Gilmour, D., Rembold, M. & Leptin, M. From morphogen to morphogenesis and back. *Nature* **541**, 311–320 (2017). <https://doi.org/10.1038/nature21348>
16. Forkert, N. D. Falling short? Mitigating equity gaps and demographic bias in AI-driven molecular diagnostics. *Expert Rev. Mol. Diagn.* **26**, 719–721 (2026).
17. Loewe, A., Hunter, P. J. & Kohl, P. Computational modelling of biological systems now and then: revisiting tools and visions from the beginning of the century. *Philos. Trans. R. Soc. A* **383**, 20230384 (2025). <https://doi.org/10.1098/rsta.2023.0384>
18. Liu, R., Pisco, A. O., Braun, E., Linnarsson, S. & Zou, J. Dynamical systems model of RNA velocity improves inference of single-cell trajectory, pseudo-time and gene regulation. *J. Mol. Biol.* **434**, 167606 (2022). <https://doi.org/10.1016/j.jmb.2022.167606>
19. Meinshausen, N. et al. Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl Acad. Sci. USA* **113**, 7361–7368 (2016). <https://doi.org/10.1073/pnas.1510493113>
20. Sadria, M. & Bury, T. M. FateNet: an integration of dynamical systems and deep learning for cell fate prediction. *Bioinformatics* **40**, btae525 (2024). <https://doi.org/10.1093/bioinformatics/btae525>
21. Dimitrov, D., Schrod, S., Rohbeck, M. et al. Interpretation, extrapolation and perturbation of single cells. *Nat Rev Genet* **27**, 349–370 (2026). <https://doi.org/10.1038/s41576-025-00920-4>
22. Li, Q.V., Dixon, G., Verma, N. et al. Genome-scale screens identify JNK–JUN signaling as a barrier for pluripotency exit and endoderm differentiation. *Nat Genet* **51**, 999–1010 (2019). <https://doi.org/10.1038/s41588-019-0408-9>
23. Song, D., Lian, Y. & Zhang, L. The potential of activator protein 1 (AP-1) in cancer targeted therapy. *Front. Immunol.* **14**, 1224892 (2023). <https://doi.org/10.3389/fimmu.2023.1224892>
24. Hu, X., Li, H., Chen, M. et al. Reference-informed evaluation of batch correction for single-cell omics data with overcorrection awareness. *Commun. Biol.* **8**, 521 (2025). <https://doi.org/10.1038/s42003-025-07947-7>

25. Fu, S., Wang, S., Si, D. et al. Benchmarking single-cell multi-modal data integrations. *Nat Methods* **22**, 2437–2448 (2025). <https://doi.org/10.1038/s41592-025-02737-9>
26. Luecken, M.D., Büttner, M., Chaichoompu, K. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* **19**, 41–50 (2022). <https://doi.org/10.1038/s41592-021-01336-8>
27. Wong, M.Y.H., Ong, J.C.L., Shah, N. et al. Neuro-symbolic artificial intelligence in medicine. *Nat. Biomed. Eng* **10**, 1507–1512 (2026). <https://doi.org/10.1038/s41551-026-01728-1>
28. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023). <https://doi.org/10.1038/s41576-023-00580-2>
29. Jackson, C. A. & Vogel, C. New horizons in the stormy sea of multimodal single-cell data integration. *Mol. Cell* **82**, 248–259 (2022). <https://doi.org/10.1016/j.molcel.2021.12.012>
30. Tejada-Lapuerta, A., Bertin, P., Bauer, S., Aliee, H., Bengio, Y. & Theis, F. J. Causal machine learning for single-cell genomics. *Nat. Genet.* **57**, 797–808 (2025). <https://doi.org/10.1038/s41588-025-02124-2>
31. Theodoris, C. V. et al. Transfer learning enables predictions in network biology. *Nature* **618**, 616–624 (2023). <https://doi.org/10.1038/s41586-023-06139-9>
32. Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 (2024). <https://doi.org/10.1038/s41592-024-02201-0>
33. Hao, M. et al. Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* **21**, 1481–1491 (2024). <https://doi.org/10.1038/s41592-024-02305-7>
34. Kedzierska, K. Z., Crawford, L., Amini, A. P. & Lu, A. X. Zero-shot evaluation reveals limitations of single-cell foundation models. *Genome Biol.* **26**, 101 (2025). <https://doi.org/10.1186/s13059-025-03574-x>
35. Ahlmann-Eltze, C., Huber, W. & Anders, S. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. *Nat. Methods* (2025). <https://doi.org/10.1038/s41592-025-02772-6>
36. Wong, D. R., Hill, A. S. & Moccia, R. Simple controls exceed best deep learning algorithms and reveal foundation model effectiveness for predicting genetic perturbations. *Bioinformatics* **41**, btaf317 (2025). <https://doi.org/10.1093/bioinformatics/btaf317>
37. Bunne, C. et al. How to build the virtual cell with artificial intelligence: priorities and opportunities. *Cell* **187**, 7045–7063 (2024). <https://doi.org/10.1016/j.cell.2024.11.015>
38. Qian, L., Dong, Z. & Guo, T. Grow AI virtual cells: three data pillars and closed-loop learning. *Cell Res.* **35**, 319–321 (2025). <https://doi.org/10.1038/s41422-025-01101-y>
39. Wei, Z. et al. Benchmarking algorithms for generalizable single-cell perturbation response prediction. *Nat. Methods* (2025). <https://doi.org/10.1038/s41592-025-02980-0>
40. Dong, M. et al. Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nat. Methods* **20**, 1769–1779 (2023). <https://doi.org/10.1038/s41592-023-02040-5>

41. Valous, N. A., Popp, F., Zornig, I., Jager, D. & Charoentong, P. Graph machine learning for integrated multi-omics analysis. *Br. J. Cancer* **131**, 205-211 (2024).
<https://doi.org/10.1038/s41416-024-02706-7>
42. Bronstein, M. M., Bruna, J., Cohen, T. & Velickovic, P. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. *Nature* **604**, 253-262 (2022). <https://doi.org/10.1038/s41586-021-04533-1>
43. Talukder, A., Barham, C., Li, X. & Hu, H. Multimodal foundation transformer models for multiscale genomics. *Nat. Methods* (2025). <https://doi.org/10.1038/s41592-025-02918-6>
44. Brynildsen, J. K., Rajan, K., Henderson, M. X. & Bassett, D. S. Network models to enhance the translational impact of cross-species studies. *Nat. Rev. Neurosci.* **24**, 575-588 (2023).
<https://doi.org/10.1038/s41583-023-00720-x>
45. Chua, M. et al. Tackling prediction uncertainty in machine learning for healthcare. *Nat. Biomed. Eng.* **7**, 711-718 (2023). <https://doi.org/10.1038/s41551-022-00988-x>
46. Zhang, J., Cammarata, L., Squires, C., Sapsis, T. P. & Uhler, C. Active learning for optimal intervention design in causal models. *Nat. Mach. Intell.* **5**, 1066-1075 (2023). <https://doi.org/10.1038/s42256-023-00719-0>
47. Wu, Y., Wershof, E., Schmon, S. et al. PerturBench: Benchmarking Machine Learning Models for Cellular Perturbation Analysis. *Adv. Neural Inf. Process. Syst.* **38** (2025).
<https://doi.org/10.52202/085713-3225>
48. Gaballa, Y., Ahmed, S. & Abdelaal, T. Benchmarking single-cell foundation models in a zero-shot setting. *bioRxiv* (2026). <https://doi.org/10.64898/2026.08.03.739553>
49. Miller, H. E., Mejia, G. M., Leblanc, F. J. A. et al. Deep learning perturbation models can outperform baselines on calibrated metrics. *Nat. Biotechnol.* (2026). <https://doi.org/10.1038/s41587-026-03307-w>
50. Adduri AK, Gautam D, Bevilacqua B, et al. Predicting cellular responses to perturbation across diverse contexts with State. *Cell*. 2026;189(19):5914-5931.e20. doi:10.1016/j.cell.2026.07.052.
51. Markov, G. J., Mai, T., Nair, S. et al. AP-1 is a temporally regulated dual gatekeeper of reprogramming to pluripotency. *Proc. Natl Acad. Sci. USA* **118**, e2104841118 (2021).
<https://doi.org/10.1073/pnas.2104841118>
52. Comandante-Lou, N., Baumann, D. G. & Fallahi-Sichani, M. AP-1 transcription factor network explains diverse patterns of cellular plasticity in melanoma cells. *Cell Rep.* **40**, 111147 (2022).
<https://doi.org/10.1016/j.celrep.2022.111147>
53. Bixi, G., Durrant, M. G., Ku, J. et al. Genome modelling and design across all domains of life with Evo 2. *Nature* **652**, 1349-1361 (2026). <https://doi.org/10.1038/s41586-026-10176-5>
54. Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C. et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Mol. Syst. Biol.* **19**, e11517 (2023).
<https://doi.org/10.15252/msb.202211517>
55. Yuan, B., Shen, C., Luna, A. et al. CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Syst.* **12**, 128-140.e4 (2021).

<https://doi.org/10.1016/j.cels.2020.11.013>

56. Roohani, Y., Huang, K. & Leskovec, J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nat. Biotechnol.* (2023). <https://doi.org/10.1038/s41587-023-01905-6>
57. Ma, J., Yu, M. K., Fong, S. et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* 15, 290-298 (2018). <https://doi.org/10.1038/nmeth.4627>

Figures

Regulation and Representation of Biology across Scales

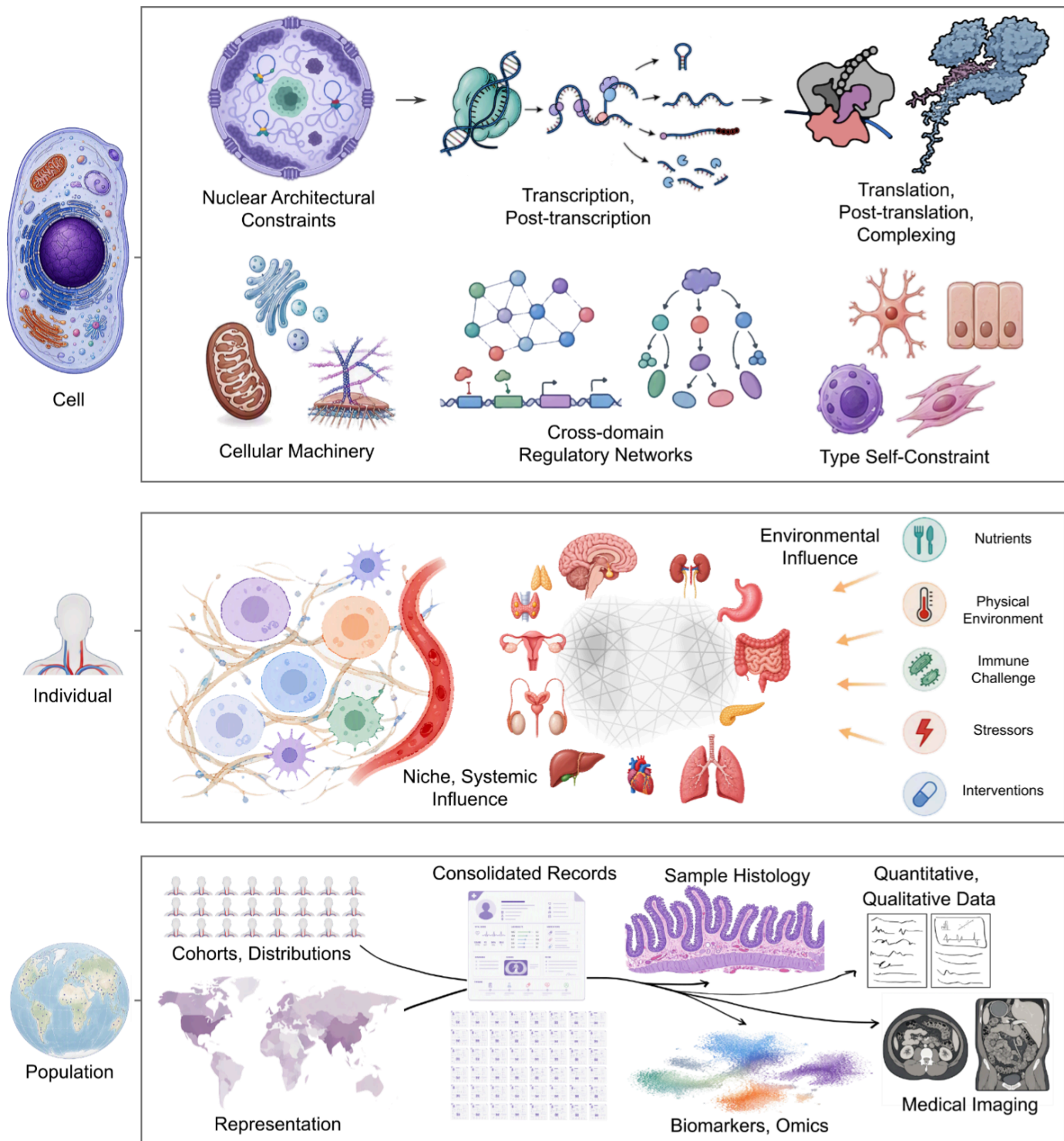


Figure 1

Biological regulation and representation across domains and scales.

Modeling Regulatory Architecture across Domains and Entities

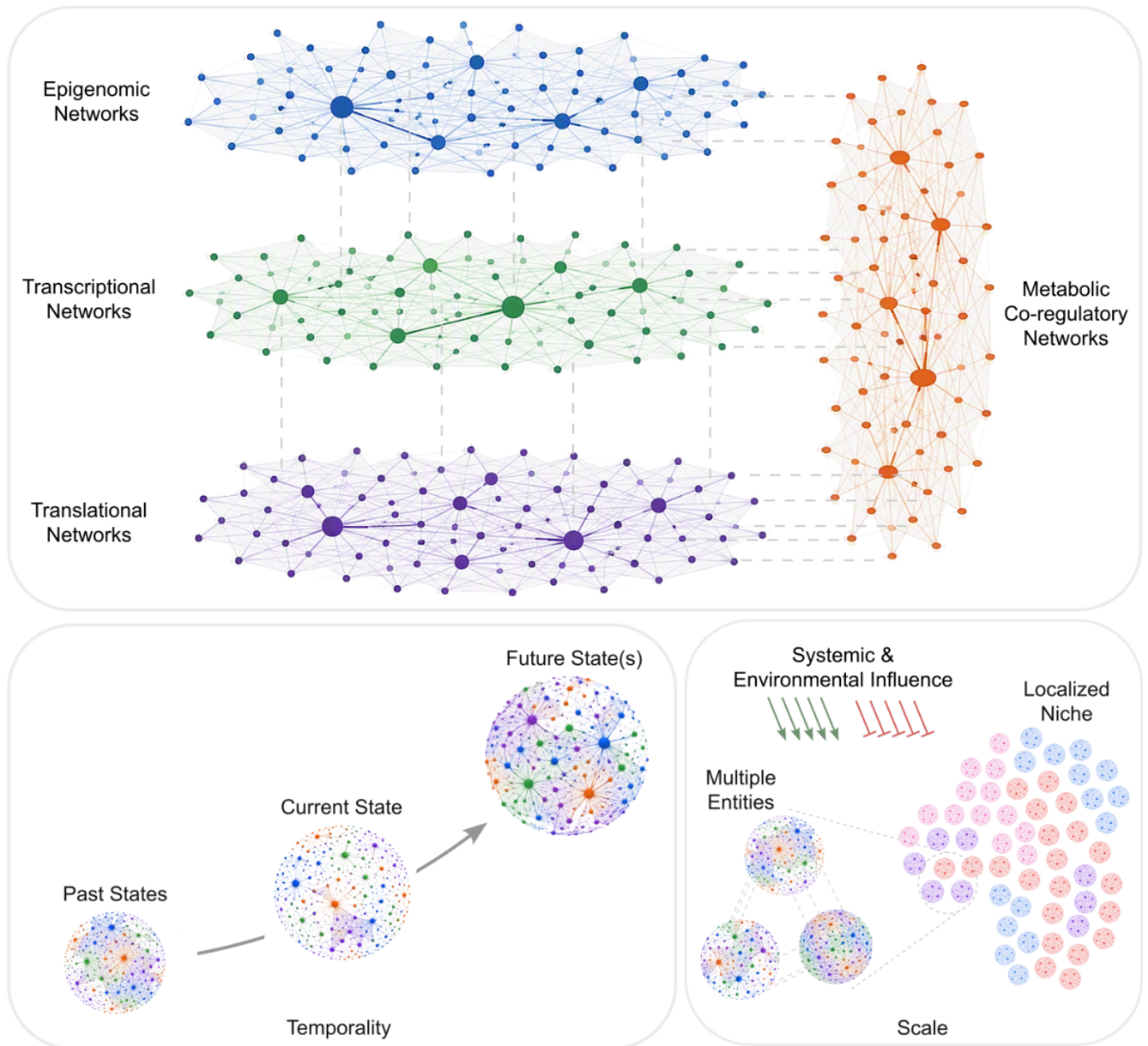


Figure 2

Conceptual multilayer regulatory architecture for modeling biological state transitions across interacting domains, entities, and timescales.

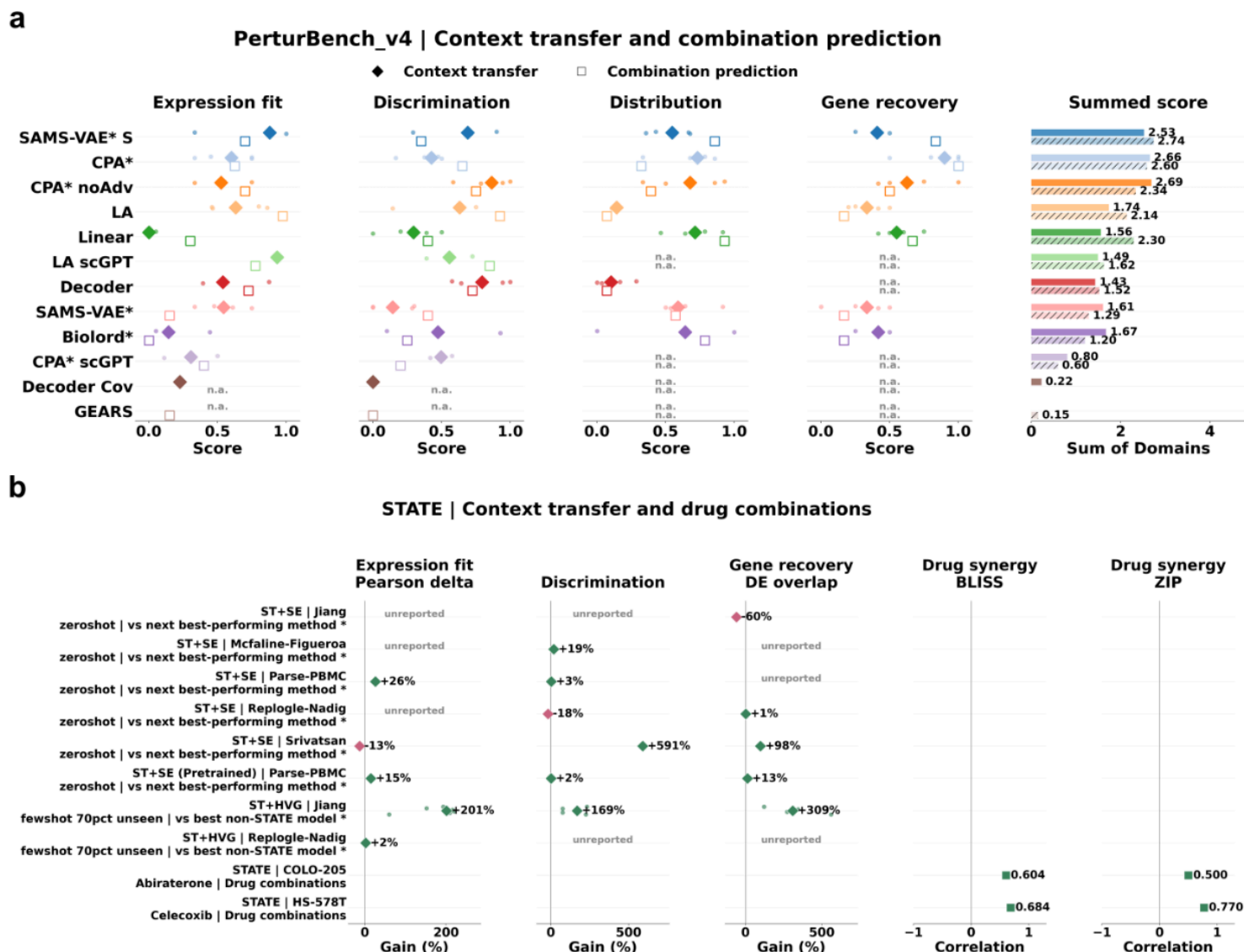


Figure 3

(a) Secondary analysis across model benchmarking literature for combination prediction and context transfer. (b) Bringing STATE into the fold. ST: State transition. SE: State embedding. HVG: Highly varying genes.