



Pharmacology Depends on Cell State When Computational Models Rank Natural Products Across Differentiation, Metabolic, Immune, and Disease-Specific Cellular Contexts

Gabriel Costa^{1*}, Lucas Ribeiro¹, Ricardo Alves², Mariana Lopes³

¹Department of Cell State-Dependent Pharmacology, Faculty of Pharmacy, Federal University of Minas Gerais, Belo Horizonte, Brazil.

²Department of Metabolic and Immune Context in Natural Product Ranking, Faculty of Pharmacy, University of Coimbra, Coimbra, Portugal.

³Department of Disease-Specific Cellular Contexts, Faculty of Pharmacy, University of São Paulo, São Paulo, Brazil.

ABSTRACT

Computational models commonly rank compounds against cell lines, cell types, or disease classes as though those labels defined a stable pharmacological context. Yet the biological state encountered by a compound can vary within the same nominal system because differentiation, metabolic conditions, inflammatory activation, disease programs, treatment history, and other processes alter the molecular configuration through which perturbations are interpreted. This article develops an original computational framework for ranking natural products as conditional compound–cell-state pairs rather than as context-free compounds. The analysis integrates evidence from perturbational transcriptomics, single-cell pharmacology, developmental and lineage biology, metabolic context, immune-state mapping, disease-specific response modeling, and recent representation-learning approaches. The proposed framework separates molecular representation from cellular-state representation, models their interaction explicitly, and treats uncertainty under unseen compounds, states, and experimental systems as part of the ranking problem rather than as a downstream qualification. It further distinguishes nominal cell identity, measured cell state, predicted molecular response, functional pharmacology, and translational relevance so that one layer is not mistaken for another. Natural-product ranking is therefore reframed around state-selective activity: a candidate can be prioritized because its predicted response is favorable in a defined state and comparatively less favorable, less certain, or mechanistically different elsewhere. The framework is not presented as prospectively validated or universally generalizable. Its value is methodological: it specifies what information a cell-state-aware ranking system should preserve, what evidence can train or test it, and which biological and computational failures must remain visible before translational interpretation.

Key Words: Cell state, Natural products, Computational pharmacology, Single-cell transcriptomics, Drug response, Molecular machine learning

eIJPPR 2025; 15(5):73-81

HOW TO CITE THIS ARTICLE: Costa G, Ribeiro L, Alves R, Lopes M. Pharmacology Depends on Cell State When Computational Models Rank Natural Products Across Differentiation, Metabolic, Immune, and Disease-Specific Cellular Contexts. Int J Pharm Phytopharmacol Res. 2025;15(5):73-81. <https://doi.org/10.51847/AmBjItCEUZ>

INTRODUCTION

A pharmacological model is often organized around a compound descriptor, a nominal cellular label, and an observed response. That arrangement is convenient, but it

can imply that “cell line,” “cell type,” or “tumor class” is a sufficiently stable biological coordinate. Cancer cell lines can undergo genetic and transcriptional evolution that changes drug response despite retaining the same nominal identity [1]. The practical consequence is not that cell lines

Corresponding author: Gabriel Costa

Address: Department of Cell State-Dependent Pharmacology, Faculty of Pharmacy, Federal University of Minas Gerais, Belo Horizonte, Brazil.

E-mail: ✉ gabriel.costa@ufmg.br

Received: 16 May 2025; **Revised:** 27 August 2025; **Accepted:** 03 September 2025

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



are unusable, but that label identity and response context are separable variables. For natural-product ranking, this distinction matters because structural novelty does not compensate for biological context that has been compressed away.

Single-cell studies sharpen the problem by showing that cellular state is neither a synonym for genotype nor a fixed subtype. Contemporary cancer-state synthesis describes recurrent transcriptional programs that coexist, change, and interact with microenvironmental influences [2]. Pan-cancer single-cell profiling similarly identifies recurring heterogeneous programs within cancer cell lines and links some of those programs to differences in drug sensitivity [3]. A model trained on a cell-line average can therefore learn a response associated with the mixture of states present during measurement rather than an invariant property of the named line. When the mixture shifts, the learned association may shift with it.

Natural products make this contextual problem especially consequential because computational prioritization already operates across chemically diverse structures, incomplete annotations, heterogeneous bioassays, and uneven mechanistic knowledge. Artificial-intelligence approaches can support natural-product discovery, including structure-based, omics-based, and data-integration tasks, but their usefulness remains dependent on evidence quality and appropriate validation [4]. The argument developed here is that biological context should be represented with comparable care. The proposed framework therefore treats pharmacological prediction as conditional: the relevant object is not simply a compound or an abstract cell, but a compound evaluated against a specified, measured, or inferable cellular state. This does not assert that state is the only determinant of response; it makes state an explicit variable whose contribution can be tested rather than silently absorbed into a label.

Cell State as a hidden variable in bioactivity prediction

Cell state becomes pharmacologically useful only when it is operationalized rather than invoked as a vague explanation for residual variability. Perturbational resources such as the L1000 Connectivity Map established that chemical exposures can be represented through reproducible transcriptional signatures at large scale [5]. Multiplexed single-cell response profiling then showed that common perturbations can yield distinct transcriptional responses across cellular backgrounds while simultaneously revealing vulnerability and mechanism-of-action information [6]. These observations justify treating the baseline cellular configuration as part of a prediction problem, but they do not establish that every response difference is state-driven; genetic background, culture history, measurement noise, dose, and time can remain competing explanations.

For this framework, cell state is proposed as a response-relevant configuration measured or inferred at the time scale appropriate to the prediction task. A state representation may encode baseline transcriptional programs, regulatory activity, pathway availability, phenotypic composition, or other experimentally defensible features, whereas nominal identity remains separate metadata. Single-cell expression profiles can already support drug-sensitivity prioritization, as illustrated by DREEP, which transfers pharmacogenomic response information to tumor single-cell profiles and validates selected predictions in independent settings [7]. The methodological step proposed here is more restrictive than calling any embedding a “state”: a representation should earn that status only if it adds reproducible response information beyond simpler identifiers and survives testing across relevant perturbations and experimental systems. The construct distinctions needed for that test are summarized in **Table 1**; they prevent nominal identity, measured state, predicted response, and downstream pharmacological interpretation from becoming interchangeable.

Table 1. Construct boundaries for cell-state-aware bioactivity prediction

Construct	Operational meaning in this framework	Appropriate computational use	Boundary
Nominal cellular identity	Cell line, cell type, or disease label	Metadata, stratification, baseline comparator	May conceal molecular drift or within-label heterogeneity [1]
Measured cellular state	Response-relevant molecular or phenotypic configuration observed before perturbation	Conditional model input	Must add reproducible information beyond the nominal label; state is not automatically causal
Perturbational signature	Molecular change measured after exposure	Training target, response descriptor, mechanism-generating evidence	Transcriptional change is not equivalent to target engagement or therapeutic effect [5]
State-resolved sensitivity estimate	Predicted response for a defined state	Candidate prioritization and comparison	Requires external validation and calibration [7]
Translational interpretation	Inference about likely usefulness beyond the assay	Hypothesis generation only	Cannot be inferred from prediction score alone

Note: Definitions identified as belonging to this framework are proposed analytical distinctions. Evidence-derived boundaries are anchored to the cited studies.

The central relationship is visualized in **Figure 1**, where compound information and a measured or inferred cellular

state remain distinct inputs before conditional response estimation.

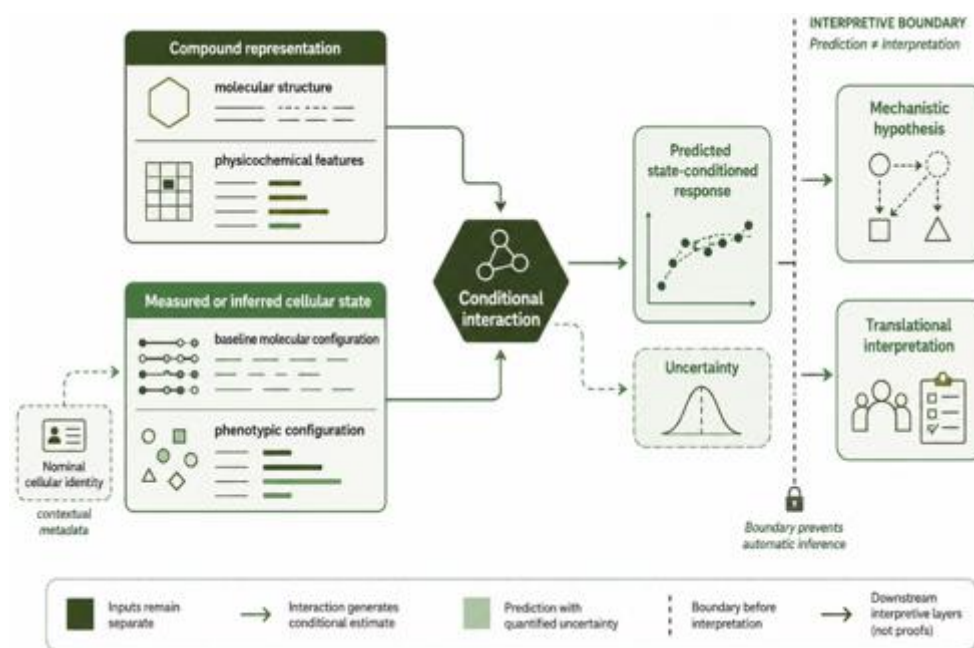


Figure 1. Conditional pharmacology begins before the response is predicted

Differentiation state changes target and pathway availability

Differentiation is a useful stress test for any claim that pharmacological response belongs to a cell label rather than to a cell in a particular configuration. In melanoma, transcriptionally defined differentiation states were associated with different vulnerability to drug-induced iron-dependent oxidative stress, providing direct evidence that lineage position can coincide with a different pharmacological liability [8]. In B-cell acute lymphoblastic leukemia, single-cell systems pharmacology linked developmental arrest to asparaginase response and identified state-associated network features that were experimentally interrogated in the relevant disease setting [9]. These studies support a conditional principle: when differentiation changes the molecular programs through which a perturbation is processed, the same compound-level description may map to a different response. They do not justify transferring the specific melanoma or leukemia mechanisms to unrelated lineages. A computational model should therefore avoid representing differentiation merely as an ordinal tag appended after prediction. The proposed approach treats lineage and differentiation information as components of the state representation when they are experimentally measurable and demonstrably response-relevant. Clone-specific therapeutic-vulnerability modeling provides a methodological precedent: CaDRReS-Sc uses single-cell transcriptomic signatures to estimate heterogeneous drug

sensitivities and to prioritize combinations across malignant subpopulations [10]. For natural products, the analogous task is not to declare a compound “active” in general, but to ask whether its predicted activity is preserved, gained, lost, or qualitatively altered as cells occupy different lineage-associated programs. This representation also permits a model to retain uncertainty when a compound is evaluated in a differentiation state that was sparse or absent during training. Such differences would still require orthogonal validation because transcriptional similarity, inferred pathway availability, and functional target engagement are distinct evidential levels.

Metabolic state changes compound response

Metabolic context demonstrates why cellular state cannot be reduced to a transcriptomic subtype alone. Screening in serum-derived medium showed that changing extracellular nutrient conditions can materially change cellular responses to compounds targeting metabolism [11]. The result is important for prediction because a model trained under one medium may attribute a response to compound structure or nominal cell identity when part of the observed effect depends on nutrient availability. Tumor-microenvironment research further shows that extracellular nutrients interact with cellular metabolic programs and can modify pathway use and metabolic dependencies [12]. These findings support metabolic state as a conditional pharmacological variable, while also

showing why it should not be represented by a single universal “metabolic score.” The relevant prediction target is the response of a defined metabolic configuration, not an assumed intrinsic sensitivity that is independent of environment.

In the proposed framework, metabolic state can be encoded through experimentally justified features such as nutrient condition, pathway activity, metabolomic measurements, or transcriptional proxies when their relationship to the prediction task has been established. The model should preserve the provenance of those measurements because culture medium, tissue environment, oxygenation, and prior treatment are not interchangeable exposures. A natural product that appears selective under one metabolic configuration should therefore receive a state-qualified prediction rather than a context-free rank. Conversely, failure to reproduce the same response under another nutrient condition need not mean that the original result was erroneous; it may expose a boundary of the pharmacological effect. This interpretation remains preclinical: an in-vitro metabolic context does not reconstruct systemic exposure, tissue distribution, metabolism, or pharmacokinetics, and state-conditioned prediction cannot bridge those levels without additional evidence.

Immune activation reconfigures pharmacological networks

Immune activation provides a different form of state dependence because the relevant configuration can be induced rapidly by extracellular signaling rather than inherited from lineage. A single-cell atlas of cytokine responses showed that the same cytokine can elicit markedly different transcriptional programs across immune-cell populations, and that cytokine combinations and cell identity jointly structure the resulting response landscape [13]. For computational pharmacology, this means that an inflammatory label such as “activated” is unlikely to be sufficient. A useful representation should specify which immune program is active, in which cellular compartment, and under which signaling context. The proposed framework therefore treats immune activation as a conditional state descriptor rather than a binary covariate. That distinction becomes pharmacologically relevant when state information changes therapeutic priorities. scDrugPrio integrates single-cell transcriptomics with disease- and drug-related network information to generate cell- and patient-specific drug rankings in immune-mediated inflammatory diseases [14]. Its results support the narrower claim that drug priority can change with disease-cell context; they do not establish that transcriptomic ranking alone predicts clinical benefit. For natural products, immune-state conditioning should

therefore be used to identify hypotheses such as selective activity in a defined inflammatory program, while preserving uncertainty about target engagement, exposure, off-target immunomodulation, and effects in other immune compartments.

Disease state alters the meaning of molecular perturbation

Disease-specific programs add another layer because cells with related lineage identities can occupy response states produced by malignant evolution, treatment pressure, or patient-specific microenvironments. PERCEPTION uses single-cell transcriptomics to predict treatment response and resistance in cancer, illustrating that patient-specific cellular programs can contribute information beyond generic drug-response averages [15]. The relevant computational lesson is not that every disease state should receive its own isolated model, but that a perturbation may have different meaning when the baseline network configuration differs.

The same principle appears at the level of inpatient clonal composition. scTherapy uses single-cell transcriptomes and molecular target information to prioritize combinations intended to co-inhibit distinct malignant clones [16]. This supports explicit representation of disease-state heterogeneity rather than averaging all malignant cells into one context. In the framework proposed here, disease programs are therefore encoded only when they are measurable and response-relevant, and their influence is estimated through the compound–state interaction rather than assumed from diagnosis alone. A prediction remains a hypothesis about a defined cellular configuration; it does not establish clinical efficacy, tolerability, pharmacokinetic exposure, or durable suppression of the disease ecosystem.

Why cell-line labels are too coarse for contextual pharmacology

Cell-line labels remain valuable identifiers, but they are too coarse to function as complete pharmacological states. Single-cell dose-response experiments show that population-level curves can conceal heterogeneous cellular responses to the same concentration of an inhibitor [17]. This measurement problem differs from the strain evolution documented for nominally identical lines [1] and from recurring transcriptional programs within a line [3]. The three phenomena can coexist: a named line may drift between laboratories, contain multiple states at one time, and show heterogeneous responses even within the same experimental condition. Treating all three as one “cell-line effect” obscures where prediction error originates.

For modeling, the implication is a hierarchy rather than a replacement. Cell-line identity can anchor provenance and experimental history; state measurements can describe the

configuration present before perturbation; dose and time can define the exposure; and post-perturbation observations can quantify response. **Table 2** distinguishes these levels because each supports a different inference and

a different failure mode. **Figure 2** then shows the proposed progression from label-only ranking toward state-qualified prediction without implying that increasing biological detail automatically guarantees better performance.

Table 2. Contextual levels that should remain distinct in computational pharmacology

Level	What it identifies	What it can legitimately support	What it can obscure
Nominal label	Named line, cell type, disease class	Provenance and stratification	Evolution within a nominal line [1]
Within-label state	Coexisting molecular program	State-specific association or ranking	Unmeasured state mixtures [3]
Exposure condition	Dose, time, medium, signaling context	Conditional comparison	Context shifts across experiments
Single-cell response	Cell-level effect distribution	Heterogeneity and responder fractions	Technical noise or transient variation [17]
State-qualified prediction	Proposed compound response for a defined state	Prioritization with explicit uncertainty	Does not by itself establish mechanism or efficacy

The resulting distinction is represented in **Figure 2** as increasing contextual resolution rather than a linear ladder of evidential certainty.

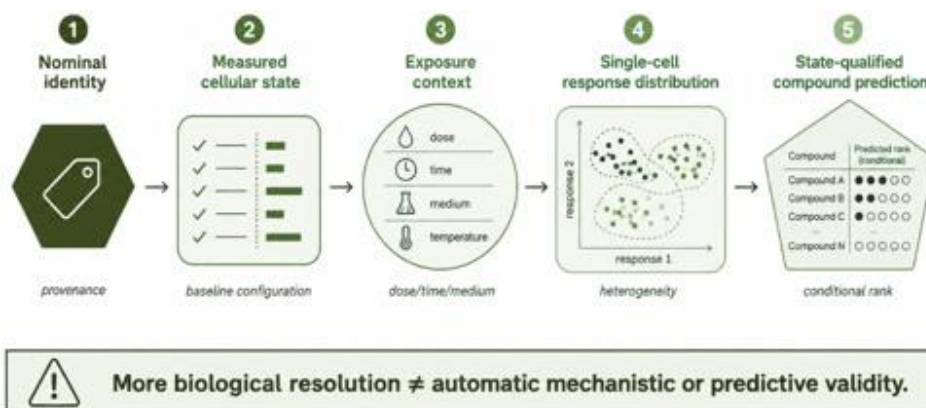


Figure 2. From label-only ranking to state-qualified pharmacological prediction

Encoding cell state alongside molecular structure

A state-aware model requires two representations that answer different questions. The compound encoder describes what perturbing chemistry is presented to the system; the state encoder describes the biological configuration that receives it. scGen established that latent representations can support prediction of single-cell perturbation responses under held-out conditions [18]. CPA extended this idea by explicitly composing basal cellular state with perturbation and dose representations [19]. These methods support conditional response modeling, but their latent variables should not be interpreted automatically as mechanistic causes.

Chemical structure can enter the same conditional logic. PRnet combines molecular information with baseline cellular expression and dose to predict transcriptional responses to novel chemical perturbations [20]. That

design is especially relevant to natural-product prioritization because it demonstrates that molecular structure and cellular context can be coupled computationally before ranking candidate responses. The present framework retains those streams as distinct encoders so that failure can be attributed, where possible, to chemical extrapolation, state extrapolation, or their interaction rather than hidden inside one fused identifier. Large pretrained models offer richer state representations but not a shortcut around validation. scGPT shows that generative pretraining across large single-cell and multi-omic corpora can produce transferable cellular representations [21]. Geneformer likewise demonstrates transfer learning for context-sensitive network predictions from large single-cell corpora [22]. Within the proposed framework these models are optional state encoders, not evidence that foundation-model embeddings are inherently

pharmacological. Any embedding must still show incremental value for the intended compound-response task, remain calibrated under state shift, and be compared with simpler representations.

A compound $\times$ cell-state conditional prediction framework

The central proposal is a conditional scoring function in which a natural product is represented by a molecular encoder, the receiving biological system by a state encoder, and their interaction by a response model that also receives experimentally relevant covariates such as dose and time. Cross-domain learning is plausible because scDEAL demonstrates that bulk pharmacogenomic supervision can be transferred to single-cell drug-response prediction [23]. The framework nevertheless retains a domain indicator and uncertainty estimate because transfer does not erase differences between bulk screens, single cells, patient-derived models, or other experimental systems.

Response should also be represented at the level demanded by the biological question. CellOT predicts perturbation-induced changes between cell-state distributions using neural optimal transport [24]. This motivates a distribution-aware option in which the output is not only a mean activity score but also a predicted shift in the composition or molecular state of the exposed population. Such a prediction would remain statistical: an optimal-transport coupling is not necessarily a true lineage trajectory, and a transcriptomic shift is not synonymous with functional benefit.

The proposed ranking function therefore has three separable outputs: expected response in the specified state, uncertainty under compound and state extrapolation, and contrast with alternative states. A candidate ranks highly only relative to the question being asked—for example, preferential perturbation of a disease-associated state while sparing a comparator state—not because it receives one universal potency score. Molecular and state encoders may be pretrained, but interaction learning and calibration must be evaluated on the actual pharmacological task. This architecture is a methodological proposal assembled from established component capabilities; no selected study prospectively validates the complete framework.

Generalizing across states without erasing biological differences

Generalization is useful only when it does not force distinct states into an apparently invariant response. Recent zero-shot evaluation of single-cell foundation models shows that pretraining alone does not guarantee strong out-of-distribution performance on biological tasks [25]. For this framework, a random train-test split is therefore insufficient. Evaluation should separate unseen compounds, unseen states, and unseen experimental

systems, because success on one axis can conceal failure on another.

Model complexity also requires a disciplined baseline. Comparative evaluation has shown that deep gene-perturbation predictors do not necessarily outperform simple linear baselines on benchmark tasks [26]. Although chemical perturbation and natural-product ranking are not identical tasks, the methodological lesson transfers: a state-aware architecture should justify every additional component against label-only, chemistry-only, state-only, and simple interaction baselines. Improvement should be assessed not only by average error but also by calibration and rank stability under biologically meaningful shifts.

The proposed generalization rule is therefore partial sharing. Common biological information may be shared across related states, while residual state-specific response should remain representable rather than penalized as noise. Distributional models provide one precedent for preserving population differences during perturbation prediction [24]. Negative transfer is itself informative: when performance deteriorates across differentiation, metabolic, inflammatory, or disease contexts, the model should expose that boundary instead of manufacturing a context-independent rank.

Single-cell measurements as training and validation evidence

Single-cell pharmacology provides the most direct experimental substrate for testing whether state-conditioned predictions correspond to heterogeneous responses. sci-Plex demonstrated massively multiplexed chemical transcriptomics at single-cell resolution across compounds, doses, and cellular systems [27]. Such designs can supply paired pre-perturbation context and post-perturbation response data at scales suitable for representation learning, although transcriptional changes remain assay readouts rather than direct evidence of therapeutic efficacy.

Validation becomes stronger when the experimental system moves beyond standard immortalized lines. A multiplex single-cell RNA-seq pharmacotranscriptomics pipeline has been applied to drug discovery in patient-derived models and can resolve drug-associated response and resistance programs [28]. State-aware ranking should therefore be tested hierarchically: within familiar systems, across held-out states, and then in biologically more relevant external models. Agreement across these levels would support transportability without implying clinical validity.

Temporal interpretation requires additional caution. scStateDynamics models drug-responsive tumor-state dynamics from pre- and post-treatment single-cell measurements and evaluates inferred transitions using

simulations and lineage-related evidence [29]. This supports using state transitions as a validation target when appropriate, but not treating every inferred path as causal

cell history. The training and validation dimensions that follow from these distinctions are summarized in **Table 3**.

Table 3. Evidence dimensions for testing a proposed cell-state-aware ranking framework

Dimension	Minimum evidence needed	Stronger test	Failure revealed
Chemical perturbation	Compound-linked single-cell response [27]	Held-out chemical structures	Chemistry extrapolation
State resolution	Baseline state measured before exposure	Held-out differentiation, metabolic, immune, or disease states	State extrapolation
Experimental system	Internal test split	Patient-derived or external model validation [28]	Domain shift
Response structure	Mean and cell-level effects	Distributional or subpopulation response	Hidden heterogeneity
Dynamics	Pre/post state measurements	Orthogonal lineage or temporal validation [29]	False trajectory inference
Model validity	Strong simple baselines	Calibration under simultaneous compound $\times$ state shift	Architectural overfitting

Note: The matrix specifies proposed validation dimensions; cited studies provide methodological precedents rather than validation of the integrated framework.

Ranking natural products for state-selective activity

Natural products add a chemical-space problem to the biological-state problem. They remain important sources of structurally and mechanistically diverse drug leads, but natural-product discovery also faces challenges involving isolation, supply, dereplication, structural characterization, target assignment, and developability [30]. A state-aware model should not treat natural origin as a favorable prior. Instead, it should ask whether a chemically verified candidate has a response profile that is desirable in the specified biological context and whether that prediction is supported inside the model's chemical and state domains. Candidate-space construction also requires provenance. COCONUT 2.0 provides a curated open collection of natural-product structures and associated metadata that can support computational candidate selection [31]. Database membership, however, is not evidence that a structure is bioactive, correctly represented in every stereochemical detail, available at the required exposure, or pharmacologically selective. Chemical identity and biological-state uncertainty should therefore remain separate components of the evidence record.

The ranking rule proposed here is comparative rather than universal. For candidate (c) in state (s), the model estimates a response distribution and uncertainty, then contrasts that estimate with predefined comparator states relevant to the intended use. PRnet demonstrates that molecular structure and cellular baseline information can support transcriptional prediction for novel chemicals [20], but the present score would add explicit cross-state contrasts and uncertainty penalties. A compound could therefore rank highly because activity is predicted to concentrate in a disease-associated or activated state while remaining weaker in a comparator state. Such state selectivity is a

prioritization hypothesis, not demonstrated therapeutic selectivity, and predictions outside supported chemical or state regions should be down-weighted rather than converted into confident ranks.

Translational boundaries and model failure

The framework can fail biologically even when its computational metrics appear acceptable. Patient-specific response models show that disease context can matter [15], while zero-shot evaluations show that representation learning can fail under distribution shift [25]. A high internal test score therefore cannot establish validity for unseen patient states, primary tissues, or natural-product classes. Likewise, heterogeneous single-cell dose responses can be hidden by aggregation [17], and additional neural complexity does not guarantee improved perturbation prediction [26].

Failure can also originate before modeling. Natural-product structures may carry provenance, stereochemical, identity, or availability uncertainties that a curated database can reduce but not eliminate [31]. On the biological side, transcriptomic state may incompletely capture metabolism, protein abundance, signaling history, microenvironment, or exposure. A state-aware model should report those missing dimensions instead of treating its embedding as a complete representation of the cell.

A further failure occurs when a state descriptor is reproducible but pharmacologically irrelevant. High-dimensional embeddings can encode donor identity, proliferation, stress, or technical variation that improves reconstruction without improving compound-response prediction. The proposed framework therefore requires ablation against simpler state descriptors and demands that any claimed state contribution improve externally tested

response prediction or uncertainty calibration, not merely representation quality.

Prospective validation must therefore separate prediction from mechanism and translation. Multiplex pharmacotranscriptomics can test whether predicted response programs appear in external experimental systems [28], but target engagement, functional phenotype, dose–time behavior, toxicity, exposure, and in-vivo relevance require additional assays. The framework is most defensible as a prioritization architecture that makes contextual assumptions and failure boundaries explicit.

CONCLUSION

Computational pharmacology becomes less interpretable when the biological recipient of a compound is represented only by a nominal label. The framework developed here instead defines natural-product ranking as a conditional prediction problem in which molecular structure, measured or inferred cellular state, exposure context, response uncertainty, and comparator states remain distinguishable. Differentiation, metabolism, immune activation, and disease programs are not proposed as interchangeable state variables; each enters only when it is measurable, pharmacologically relevant, and supported at the appropriate evidential level.

The principal contribution is therefore architectural rather than predictive. A candidate should be ranked for what it is expected to do in a specified cellular configuration, with uncertainty increasing when either chemistry or biology leaves the supported domain. State-aware prediction must also remain separate from target engagement, mechanism, efficacy, safety, and clinical usefulness. The framework can be rejected or revised if richer state information fails to improve calibrated prediction over simpler baselines, if apparent state selectivity does not reproduce experimentally, or if transfer across systems erases biologically meaningful differences. Its purpose is to make those tests possible.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*. 2018;560(7718):325-30. doi:10.1038/s41586-018-0409-3
- [2] Tirosh I, Suvà ML. Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors. *Cancer Cell*. 2024;42(9):1497-506. doi:10.1016/j.ccell.2024.08.005
- [3] Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet*. 2020;52(11):1208-18. doi:10.1038/s41588-020-00726-6
- [4] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov*. 2023;22(11):895-916. doi:10.1038/s41573-023-00774-7
- [5] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437-52.e17. doi:10.1016/j.cell.2017.10.049
- [6] McFarland JM, Paoletta BR, Warren A, Geiger-Schuller K, Shibue T, Rothberg M, et al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat Commun*. 2020;11(1):4296. doi:10.1038/s41467-020-17440-w
- [7] Pellicchia S, Viscido G, Franchini M, Gambardella G. Predicting drug response from single-cell expression profiles of tumours. *BMC Med*. 2023;21(1):476. doi:10.1186/s12916-023-03182-1
- [8] Tsoi J, Robert L, Paraiso KHT, Galvan C, Sheu KM, Lay J, et al. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell*. 2018;33(5):890-904.e5. doi:10.1016/j.ccell.2018.03.017
- [9] Huang X, Li Y, Zhang J, Yan L, Zhao H, Ding L, et al. Single-cell systems pharmacology identifies development-driven drug response and combination therapy in B cell acute lymphoblastic leukemia. *Cancer Cell*. 2024;42(4):552-67.e6. doi:10.1016/j.ccell.2024.03.003
- [10] Suphavitai C, Chia S, Sharma A, Tu L, Peres Da Silva R, Mongia A, et al. Predicting heterogeneity in clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. *Genome Med*. 2021;13(1):189. doi:10.1186/s13073-021-01000-y
- [11] Abbott KL, Ali A, Casalena D, Do BT, Ferreira R, Cheah JH, et al. Screening in serum-derived medium reveals differential response to compounds targeting metabolism. *Cell Chem Biol*. 2023;30(9):1156-68.e7. doi:10.1016/j.chembiol.2023.08.007

- [12] Lobel GP, Jiang Y, Simon MC. Tumor microenvironmental nutrients, cellular responses, and cancer. *Cell Chem Biol*. 2023;30(9):1015-32. doi:10.1016/j.chembiol.2023.08.011
- [13] Cui A, Huang T, Li S, Ma A, Pérez JL, Sander C, et al. Dictionary of immune responses to cytokines at single-cell resolution. *Nature*. 2024;625(7994):377-84. doi:10.1038/s41586-023-06816-9
- [14] Schäfer S, Smelik M, Sysoev O, Zhao Y, Eklund D, Lilja S, et al. scDrugPrio: A framework for the analysis of single-cell transcriptomics to address multiple problems in precision medicine in immune-mediated inflammatory diseases. *Genome Med*. 2024;16(1):42. doi:10.1186/s13073-024-01314-7
- [15] Sinha S, Vegesna R, Mukherjee S, Kammula AV, Dhruva SR, Wu W, et al. PERCEPTION predicts patient response and resistance to treatment using single-cell transcriptomics of their tumors. *Nat Cancer*. 2024;5(6):938-52. doi:10.1038/s43018-024-00756-7
- [16] Ianevski A, Nader K, Driva K, Senkowski W, Bulanov D, Moyano-Galceran L, et al. Single-cell transcriptomes identify patient-tailored therapies for selective co-inhibition of cancer clones. *Nat Commun*. 2024;15(1):8579. doi:10.1038/s41467-024-52980-5
- [17] Kinnunen PC, Humphries BA, Luker GD, Luker KE, Linderman JJ. Characterizing heterogeneous single-cell dose responses computationally and experimentally using threshold inhibition surfaces and dose-titration assays. *NPJ Syst Biol Appl*. 2024;10(1):42. doi:10.1038/s41540-024-00369-x
- [18] Lotfollahi M, Wolf FA, Theis FJ. scGen predicts single-cell perturbation responses. *Nat Methods*. 2019;16(8):715-21. doi:10.1038/s41592-019-0494-8
- [19] Lotfollahi M, Klimovskaia Susmelj A, De Donno C, Hetzel L, Ji Y, Ibarra IL, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Mol Syst Biol*. 2023;19(6). doi:10.15252/msb.202211517
- [20] Qi X, Zhao L, Tian C, Li Y, Chen ZL, Huo P, et al. Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nat Commun*. 2024;15(1):9256. doi:10.1038/s41467-024-53457-1
- [21] Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods*. 2024;21(8):1470-80. doi:10.1038/s41592-024-02201-0
- [22] Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, et al. Transfer learning enables predictions in network biology. *Nature*. 2023;618(7965):616-24. doi:10.1038/s41586-023-06139-9
- [23] Chen J, Wang X, Ma A, Wang QE, Liu B, Li L, et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. *Nat Commun*. 2022;13(1):6494. doi:10.1038/s41467-022-34277-7
- [24] Bunne C, Stark SG, Gut G, Sarabia del Castillo J, Levesque M, Lehmann KV, et al. Learning single-cell perturbation responses using neural optimal transport. *Nat Methods*. 2023;20(11):1759-68. doi:10.1038/s41592-023-01969-x
- [25] Kedzierska KZ, Crawford L, Amini AP, Lu AX. Zero-shot evaluation reveals limitations of single-cell foundation models. *Genome Biol*. 2025;26(1):101. doi:10.1186/s13059-025-03574-x
- [26] Ahlmann-Eltze C, Huber W, Anders S. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. *Nat Methods*. 2025;22(8):1657-61. doi:10.1038/s41592-025-02772-6
- [27] Srivatsan SR, McFaline-Figueroa JL, Ramani V, Saunders L, Cao J, Packer J, et al. Massively multiplex chemical transcriptomics at single-cell resolution. *Science*. 2020;367(6473):45-51. doi:10.1126/science.aax6234
- [28] Dini A, Barker H, Piki E, Sharma S, Raivola J, Murumägi A, et al. A multiplex single-cell RNA-Seq pharmacotranscriptomics pipeline for drug discovery. *Nat Chem Biol*. 2025;21(3):432-42. doi:10.1038/s41589-024-01761-8
- [29] Guo W, Li X, Wang D, Yan N, Hu Q, Yang F, et al. scStateDynamics: Deciphering the drug-responsive tumor cell state dynamics by modeling single-cell level expression changes. *Genome Biol*. 2024;25(1):297. doi:10.1186/s13059-024-03436-y
- [30] Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. Natural products in drug discovery: Advances and opportunities. *Nat Rev Drug Discov*. 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z
- [31] Chandrasekhar V, Rajan K, Kanakam SRS, Sharma N, Weissenborn V, Schaub J, et al. COCONUT 2.0: A comprehensive overhaul and curation of the collection of open natural products database. *Nucleic Acids Res*. 2025;53(D1). doi:10.1093/nar/gkae1063