



Single-Cell Perturbation Signatures Can Resolve Heterogeneous Natural Product Responses Across Cell States, Subpopulations, Pathways, and Therapeutic Sensitivities

Ethan Wright^{1*}, Chloe Bennett², Jack Turner¹, Olivia Harris³

¹Department of Single-Cell Perturbation Signatures, Faculty of Pharmacy, University of Leeds, Leeds, United Kingdom.

²Department of Heterogeneous Natural Product Responses, Faculty of Pharmacy, University of Sheffield, Sheffield, United Kingdom.

³Department of Therapeutic Sensitivity and Pathway Resolution, Faculty of Pharmacy, University of York, York, United Kingdom.

ABSTRACT

Population-averaged pharmacology can compress biologically different cellular responses into the same apparent treatment effect, obscuring whether a perturbation acts broadly, selectively, incompletely, or only within particular pre-existing states. This problem is especially consequential for chemically diverse natural products, whose cellular effects cannot be assumed to be homogeneous across complex populations. This original computational framework proposes the single-cell perturbation signature as a structured pharmacological response representation that preserves baseline cellular state, perturbation-induced change, subpopulation-level response, pathway-level organization, chemical context, and uncertainty rather than reducing response to a single population statistic. The framework separates measured response from computationally inferred structure and distinguishes association with baseline state from causal state dependence. It further treats response magnitude as potentially continuous, prevents resistance from being inferred solely from persistence or incomplete response, and uses pathway decomposition as an interpretive layer rather than mechanistic proof. The proposed architecture is intended to support later modeling of compound–cell-state interactions, chemical structure, response trajectories, and breadth versus selectivity while retaining explicit boundaries around prediction, mechanism, and therapeutic interpretation. Its methodological value lies in organizing heterogeneous pharmacological information without requiring every cell to conform to a single responder class. Prospective evaluation will require perturbation-specific validation across biological contexts, assays, and chemical domains. Major limitations include confounding, sparse perturbation coverage, unstable state definitions, measurement noise, model dependence, and uncertain transportability. The contribution is therefore a testable computational framework, not a validated causal or clinical model.

Key Words: Single-cell pharmacology, Perturbation signatures, Natural products, Cell-state heterogeneity, Drug response, Computational pharmacology

eIJPPR 2026; 16(2):195-203

HOW TO CITE THIS ARTICLE: Wright E, Bennett C, Turner J, Harris O. Single-Cell Perturbation Signatures Can Resolve Heterogeneous Natural Product Responses Across Cell States, Subpopulations, Pathways, and Therapeutic Sensitivities. Int J Pharm Phytopharmacol Res. 2026;16(2):195-203. <https://doi.org/10.51847/SHX3AaDFDx>

INTRODUCTION

Natural products remain an important source of pharmacologically active matter and span substantial chemical and biological diversity [1]. That diversity,

however, does not justify assuming that natural products are uniformly multitarget, intrinsically state selective, or superior to other chemical classes. The more immediate problem is representational: a population-level response can arise from many different cellular configurations. A

Corresponding author: Ethan Wright

Address: Department of Single-Cell Perturbation Signatures, Faculty of Pharmacy, University of Leeds, Leeds, United Kingdom..

E-mail: ethan.wright@leeds.ac.uk

Received: 26 January 2025; **Revised:** 20 April 2026; **Accepted:** 23 April 2026

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.



modest bulk effect could reflect a weak response in most cells, a strong response restricted to a minority, opposing changes in different subpopulations, or a shift in population composition. These possibilities are pharmacologically nonequivalent even when their population averages are similar.

Single-cell chemical perturbation experiments demonstrate that drug-induced transcriptional responses can be resolved as distributions across individual cells rather than only as averaged profiles [2, 3]. The resulting heterogeneity should not automatically be interpreted as mechanism, target engagement, or therapeutic significance, but it changes what must be represented if the analytical objective is to understand which cells respond. This article therefore develops an Original computational framework in which a perturbation is represented conditionally on cellular context and resolved across response magnitude, subpopulation structure, pathway organization, and uncertainty. The framework is proposed as an analytical architecture: it organizes evidence-bearing distinctions and generates testable relationships, but it is not presented as prospectively validated or causally established.

Natural product pharmacology across heterogeneous cell populations

Heterogeneous pharmacology begins before a compound is added. A mixed cellular system can contain distinct baseline states, genetic subclones, differentiation programs, stress states, and variable proportions of cell

types. In patient-derived cancer systems, distinct malignant clones have shown different pharmacological vulnerabilities under tested treatment conditions [4]. Such evidence establishes that coexisting cellular populations can differ in drug sensitivity, but it does not establish that every natural product produces state-selective effects or that an ex vivo clone-specific response predicts therapeutic effectiveness. For natural-product perturbation, the appropriate default is therefore conditional rather than universal: chemical exposure acts on a pre-existing biological mixture whose composition and state may shape the observed response.

Several sources of apparent heterogeneity must consequently remain separate. Heterogeneous exposure or measurement can create unequal readouts without true biological sensitivity differences; baseline-state variation exists before treatment; differential sensitivity concerns unequal perturbational response conditional on exposure; resistance implies persistence of function or viability despite a relevant challenge and should not be inferred from a single descriptive state; adaptation denotes an induced change over time; and compositional change alters the relative abundance of cellular groups. These distinctions matter because a bulk endpoint can collapse all of them into one number. **Table 1** formalizes the manuscript-specific constructs that must be kept analytically distinct before a computational response model is introduced.

Table 1. Construct boundaries required before modeling heterogeneous natural-product response

Construct	Operational meaning in this framework	What it must not be collapsed into	Interpretive status
Population-average response	Aggregate change across the sampled population	Uniform response of every cell	Measured summary
Baseline cellular state	Pre-perturbation molecular or phenotypic context	Drug-induced state change	Measured or estimated context
Perturbation-induced change	Difference associated with exposure relative to baseline or control	Pre-existing state identity	Measured or modeled response
Subpopulation-specific response	Response concentrated within a cellular subset	Whole-population potency	Evidence-supported analytical distinction
Heterogeneous exposure/readout	Unequal effective exposure or measurement across cells	Biological sensitivity	Potential confounder
Persistence, adaptation, resistance	Distinct post-perturbation interpretations requiring different evidence	A single “non-responder” class	Proposed evidence-gated distinction

Note: The final row is a proposed interpretive distinction. Persistence or incomplete change alone is insufficient to establish resistance or adaptation.

Defining a single-cell perturbation signature

A single-cell perturbation signature is proposed here as a conditional representation of how a cellular population is distributed before and after a defined perturbation, together with the context needed to interpret that change. Harmonized perturbation resources show why retaining distributional information and experimental metadata matters when responses vary across cells and datasets [5]. The proposed signature therefore differs from a bulk

differential-expression vector: it does not assume that the mean shift adequately represents the population. It also differs from a cell-type label, because cells sharing an annotation may respond differently; from a single pathway score, because multiple programs and compositional changes may coexist; and from a generic embedding, because the representation must retain pharmacologically interpretable conditioning variables.



Response strength should also remain continuous when the data support continuity. Single-cell perturbation studies have demonstrated substantial within-perturbation variation, including partial responses [6]. Independently, local perturbation effects can be quantified continuously across cellular state space without relying exclusively on predefined clusters [7]. These findings support preserving cell-level response magnitude before assigning categorical labels. The manuscript therefore proposes that a usable signature contain, at minimum when measured or estimable, baseline-state information, perturbation context, a cell-resolved response distribution, subpopulation structure, and uncertainty; dose, time, pathway programs, and chemical representation are added when available. This definition is an original integration of established analytical components and remains to be prospectively tested as a pharmacological representation.

Separating baseline cell state from drug-induced state change

The central computational problem is not merely to identify post-treatment clusters but to distinguish what was already present from what changed after perturbation. Disentangled single-cell models show that known attributes and latent variation can be represented separately under explicit modeling assumptions [8]. Perturbation-prediction models likewise demonstrate that baseline cellular states can be related to estimated post-perturbation shifts across contexts [9]. In this framework, baseline state is therefore represented as a pre-perturbation condition on response, whereas induced change is represented as a conditional displacement or distributional difference. Keeping them separate prevents a state that is common among responders from being mislabeled as a drug-created state simply because it is enriched after exposure.

Computational separation does not establish causal state dependence. In unpaired single-cell experiments, baseline composition, hidden covariates, selection effects, and treatment-associated changes can be confounded; causal identification therefore requires explicit assumptions and methods designed to address such structure [10]. A cell state associated with a stronger perturbational shift may be predictive without being the cause of sensitivity. Conversely, post-treatment enrichment of a state may reflect preferential survival, state transition, or measurement effects. The proposed framework accordingly treats baseline-response associations as a distinct evidence level from causal explanations. A causal claim would require a design capable of intervening on, or otherwise identifying, the relevant state variable rather than relying only on observational separation.

Identifying responsive and resistant subpopulations

Subpopulation analysis should resolve heterogeneous pharmacology without forcing every cell into a binary sensitive-versus-resistant scheme. Untreated single-cell state information has been used to infer drug-responsive cell types in evaluated cancer contexts, and tumor single-cell transcriptomes have also supported prediction of treatment response and resistance in tested settings [11, 12]. These studies support the value of state-informed stratification, but neither baseline association nor predictive classification alone identifies a causal resistance mechanism. In the proposed framework, a responsive subpopulation is therefore defined descriptively by a reproducible perturbation-associated change, while partial response represents reduced or incomplete change relative to the responsive distribution. Pre-existing state differences remain covariates until evidence shows they modify response.

The terms persistence, tolerance, adaptation, and resistance require stronger separation. Persistence can describe continued presence of a state or phenotype after treatment; adaptation implies an induced transition or altered program over time; tolerance usually requires evidence of maintained viability or function under exposure; and resistance requires a defensible pharmacological criterion showing failure of an otherwise relevant perturbation. A single post-treatment transcriptomic snapshot may be consistent with any of these but cannot, by itself, discriminate them. The framework therefore treats categorical resistance labels as evidence-gated interpretations layered above cell-resolved response distributions, with temporal, functional, or orthogonal pharmacological evidence required whenever the claim moves beyond descriptive persistence.

Pathway-level interpretation of heterogeneous responses

Gene-level response distributions can become difficult to interpret when different cells activate overlapping programs with different magnitudes. Large single-cell perturbation screens have shown that pathway-scale molecular signatures can be reconstructed across cellular contexts under direct perturbation designs [13]. This supports using pathway or gene-program structure as an intermediate analytical layer between individual genes and whole-cell response classes. For natural-product pharmacology, such decomposition can reveal whether two subpopulations that appear similarly responsive at an aggregate level differ in the transcriptional programs through which that response is expressed. It can also distinguish a localized pathway-associated response from a broadly distributed stress program, provided that the analysis remains conditional on cell state and perturbation context.

Pathway-level interpretation nevertheless has a strict evidentiary boundary. Enrichment, inferred activity, transcriptional signature, and functional dependence are not interchangeable. A pathway score may summarize coordinated expression without showing that the pathway is necessary for the observed phenotype, and a treatment-associated program may arise indirectly through composition or downstream consequences. Condition-associated differences can also reflect changes in the abundance of cellular neighborhoods rather than within-cell regulation [14]. The proposed response-map therefore keeps compositional change and intracellular pathway response as separate channels. Mechanistic claims require additional intervention or orthogonal functional evidence; pathway decomposition is used to organize heterogeneous response, not to convert association into demonstrated pathway dependence.

Learning compound–cell-state interactions

The learning problem is not to predict “drug response” in the abstract, but to estimate a response distribution conditional on a chemically specified perturbation and a biologically specified cellular context. Deep generative models have shown that transcriptional responses to previously unseen chemical perturbations can be modeled from compound and cellular information within evaluated domains [15]. Compositional perturbation models further show that perturbation identity, dose, and cellular

covariates can be encoded as distinct predictive factors [16]. The proposed framework adopts this separation operationally but does not assume that these factors are biologically independent. An interaction function must be allowed to represent non-additive dependence between chemical representation and baseline state, while dose, time, assay setting, and other contextual variables remain explicit rather than being absorbed silently into a latent embedding.

A second issue is the response object itself. Because control and perturbed single-cell measurements are generally unpaired, neural optimal-transport approaches demonstrate the value of modeling mappings between cell distributions rather than pretending that each observed post-treatment cell has a known pre-treatment counterpart [17]. Biological inductive priors can also support cell-type-specific drug-response prediction in small-data regimes [18]. These precedents motivate a proposed compound–cell-state interaction representation that combines a chemical descriptor or learned molecular representation, a baseline-state representation, perturbation context, and a distributional response target. Priors may constrain this interaction when evidence is sparse, but they require sensitivity analysis because an incorrect prior can impose structure that the data do not support. **Table 2** separates the computational objects from the interpretations that should not be attached to them automatically.

Table 2. Computational distinctions required for learning compound–cell-state response

Dimension	Representation in the proposed framework	Legitimate analytical use	Interpretation that remains separate
Chemical perturbation	Molecular descriptor, learned structure representation, or defined compound identity	Condition response on chemical information	Pharmacological equivalence or target engagement
Baseline cell state	Pre-perturbation molecular or phenotypic representation	Condition expected response on cellular context	Causal determinant of sensitivity
Dose, time, assay context	Explicit covariates when available	Define the perturbation condition	Interchangeable nuisance variables
Interaction function	Proposed nonlinear relation between chemical and state representations	Learn context-dependent response	Proven molecular mechanism
Response object	Distribution of induced cellular states or response magnitudes	Preserve heterogeneous outcomes	One-to-one cell trajectory
Biological prior	Optional pathway, target, or cell-type information	Regularize learning under limited data	Ground truth when inconsistent with observations

Note: The interaction function and its integrated role are proposed components. Model output remains computational prediction until experimentally observed.

A single-cell response-map framework for natural products

Perturbation analysis can be organized as a modular sequence of preprocessing, response quantification, comparison, and interpretation [19]. Building on that modular logic, this article proposes a Single-Cell Response-Map Framework in which the fundamental object is not a scalar compound effect but a context-conditioned map linking chemical perturbation to

distributions of cellular change. The map receives two information classes that remain separate: chemical context, including structure or compound identity and available dose/time information; and biological context, including the baseline distribution of cell states and relevant sample or disease context. A proposed interaction core relates these inputs without presuming that either chemical similarity or baseline-state similarity guarantees equivalent response.

The core resolves the perturbation into several analytically separate outputs. First, induced state change is represented relative to baseline rather than as post-treatment identity alone. Second, response magnitude is preserved as a distribution, permitting localized, partial, or broad effects. Third, subpopulation structure records where response is concentrated and whether composition shifts. Fourth, pathway programs summarize coordinated response without being promoted automatically to mechanism. Temporal information, when available, can organize trajectories across repeated measurements; otherwise inferred transport or pseudotime remains explicitly model-based. Finally, a proposed breadth-selectivity layer describes how widely response is distributed across states without ranking therapeutic value. Computational uncertainty, biological uncertainty, and external validation are not downstream afterthoughts but boundaries on every interpretation.

Operationally, the map preserves three evidence statuses. Observed layers contain measured baseline and post-perturbation data. Inferred layers contain model-derived state changes, distributions, pathway programs, or

trajectories whose validity depends on assumptions and available controls. Proposed interpretive layers connect those objects into response breadth, state selectivity, and validation decisions. Keeping these statuses separate is essential because the same graphical or latent representation can otherwise make predicted relationships appear experimentally observed. The framework therefore requires every output to retain its provenance from measurement through transformation to interpretation, with uncertainty propagated rather than discarded at the transition between levels.

The framework is therefore a representation-and-inference architecture, not a validated causal model. Its outputs may include measured single-cell responses, model-estimated state changes, inferred pathway programs, and proposed summary interpretations; these outputs carry different evidence status. **Figure 1** makes the critical topology visible: separate chemical and baseline-state lanes converge on a proposed interaction core, but the resolved response then diverges across subpopulation, pathway, trajectory, and breadth-selectivity structures while uncertainty and external validation constrain every level.

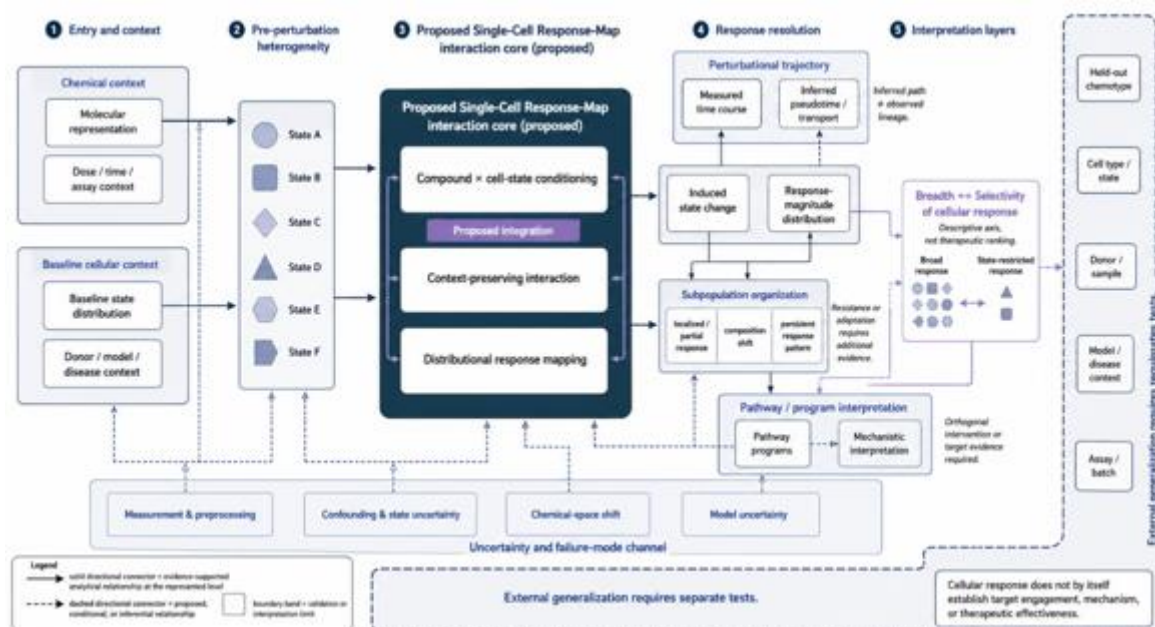


Figure 1. Resolving a natural-product perturbation from chemical and baseline context into bounded single-cell response maps

Connecting chemical structure to subpopulation-specific effects

Chemical structure offers a route from compound identity to generalizable representation, but only if the representation is evaluated against the response task. Recent drug-perturbation models show that chemical and cellular representations can be learned jointly [20], while conditional distribution models can preserve perturbation,

dose, and cell-context information in the response mapping [21]. This supports a structure-conditioned response model in which a molecule is not mapped directly to one expected phenotype; instead, its representation participates in an interaction with baseline state and context. The resulting output can therefore differ across subpopulations even when the chemical input is unchanged.

For chemical modeling, this means that a molecular representation is informative only relative to the domain in which its response relationship was learned. A descriptor may preserve stereochemistry, topology, physicochemical attributes, or learned substructural features, yet none guarantees that the biologically decisive feature has been encoded. The response map therefore treats chemical representation choice as a testable modeling decision rather than a chemically neutral preprocessing step.

Prediction for unseen small molecules provides stronger evidence that structure can carry transferable response information within evaluated chemical domains [22]. Structure and basal cellular information have also been used to learn induced transcriptional representations [23]. Neither finding licenses the inference that neighboring structures are pharmacologically equivalent, that an embedding has discovered a molecular target, or that performance transfers automatically to natural-product chemical space. Natural-product compounds may fall outside the chemical domain represented during training. The framework therefore treats chemical-domain coverage as an explicit validation variable: structural extrapolation is a prediction problem whose uncertainty should increase when chemotype support weakens, and any claimed subpopulation-specific activity still requires empirical perturbation evidence.

Discovering mechanism from perturbational trajectories

A static post-treatment profile can reveal heterogeneous response but often cannot distinguish transient stress from stable adaptation, delayed response, recovery, or selection of a pre-existing state. Trajectory analysis can compare dynamic single-cell programs and align gene-level changes across inferred cellular paths [24]. Within the proposed response-map, measured longitudinal sampling is therefore the strongest trajectory representation, followed by repeated cross-sectional time points, while pseudotime and model-implied transport paths are treated as inferential constructs. The latter may organize response states coherently, but they are not observed histories of individual cells.

Mechanistic interpretation requires an additional evidentiary step. A trajectory in which a pathway program precedes a phenotypic change may generate a mechanistic hypothesis, yet ordering alone does not demonstrate necessity, sufficiency, direct target engagement, or causal mediation. Stronger support would require perturbing the implicated pathway or target, measuring target engagement where appropriate, reproducing the temporal sequence, and linking molecular change to a functional outcome. Thus, trajectories are valuable because they convert response from a snapshot into an ordered hypothesis space; they do not by themselves identify the molecular mechanism of a natural product.

Prioritizing compounds by breadth versus selectivity of cellular response

A population average provides little information about how widely a perturbation is distributed across cell states. Because response magnitude can vary continuously among perturbed cells [6], the framework proposes breadth as the extent to which a reproducible response spans the baseline state distribution and selectivity as concentration of response within particular states or subpopulations. Evidence that baseline single-cell states can associate with drug-responsive cell types [11] supports the relevance of subset-specific response, but it does not establish a universal metric for selectivity. These concepts are therefore descriptive axes rather than numerical scores.

Breadth is not intrinsically desirable: a broad response may represent coordinated action across disease-relevant states, but it may also reflect generalized stress or nonspecific toxicity. Selectivity is equally ambiguous: restriction to a subpopulation may identify a useful vulnerability, insufficient exposure, a rare permissive state, or an assay artifact. The proposed framework consequently interprets breadth and selectivity jointly with response magnitude, pathway organization, temporal persistence, and biological context. **Table 3** organizes these dimensions as questions to be evaluated rather than as weights in a compound-ranking algorithm.

Table 3. Proposed framework dimensions for interpreting breadth and selectivity without a scoring system

Dimension	Interpretive question	Distinction preserved	Validation need
Response coverage	Across how much of the baseline state distribution is reproducible change observed?	Broad distribution versus state-restricted response	Replicate perturbation across relevant states
Response magnitude	How strong is change within responding cells?	Weak widespread change versus strong localized change	Orthogonal response measurement
Subpopulation concentration	Is response enriched in a defined cellular subset?	Selectivity versus population composition	Confirm subset identity and exposure
Temporal persistence	Does the pattern persist, adapt, recover, or emerge later?	Snapshot response versus trajectory	Dose/time-resolved perturbation
Pathway concordance	Are response programs shared or state specific?	Common program versus heterogeneous programs	Functional pathway testing where mechanism is claimed

Context stability	Does the pattern recur across biological contexts?	Within-context pattern versus transportable response	External donor/model/disease validation
-------------------	--	--	---

Note: These are proposed descriptive dimensions, not scores, ranks, therapeutic-value weights, or efficacy thresholds.

Validation across donors, models, and disease states

Validation should test the specific form of generalization being claimed. Prediction in an unseen cell type and transfer across experimental platforms are distinct challenges and can be evaluated separately [25]. Accordingly, the proposed framework distinguishes held-out chemical structure, cell state or cell type, donor or sample, model system, disease context, assay platform, perturbation batch, and combinations of these. A random within-dataset split is useful for model development but cannot establish transport across biological contexts that were represented during training.

Validation should also distinguish reproducibility from transportability. Reproducing a response within the same donor, model, assay, and chemotype neighborhood supports measurement stability; reproducing it after changing one of those axes tests a stronger claim. Changing several simultaneously tests a still harder composite shift and makes failure harder to localize. The proposed validation logic therefore favors staged external tests so that chemical-domain failure, biological-context failure, and assay-platform failure can be diagnosed separately.

Evaluation design can itself create optimistic results. Systematic variation unrelated to the perturbation-specific signal can contribute to apparent prediction performance [26]. Complex models also require comparison with strong simple baselines; deep-learning approaches have not consistently outperformed simple linear alternatives in evaluated genetic perturbation tasks [27]. These findings do not determine which model family will perform best for natural-product pharmacology, but they justify a validation hierarchy in which model complexity earns interpretation only after perturbation-specific signal and appropriate baselines are demonstrated.

Biological replication adds another level. Sample-level heterogeneity should be represented separately from cell-level heterogeneity when multiple specimens or donors are analyzed [28]. Broad benchmarking further shows that perturbation-response generalization depends on the held-out scenario rather than constituting a single model property [29]. Prospective evaluation of the response-map should therefore hold out relevant donors, disease states, model systems, chemotypes, and assay conditions rather than extrapolating one successful transfer result to all of them. Direct intervention maps demonstrate the value of experimentally perturbing biological systems when validating state–response relations [30]; for natural products, analogous validation would require controlled

chemical exposure plus orthogonal functional or mechanistic assays appropriate to the claim.

Computational and biological limits

The framework inherits limitations from both single-cell measurement and perturbation modeling. Integration and batch-correction methods trade removal of technical structure against preservation of biological variation [31]. A response map can therefore be distorted if preprocessing removes state-dependent pharmacological signal or preserves batch structure that resembles response. Missing dose or time information further makes nominally identical perturbations non-equivalent, while sparse coverage can leave particular baseline states, chemotypes, or response trajectories unsupported. Cell-state annotations are also model- and resolution-dependent; the framework consequently treats state identity as an estimated contextual variable rather than immutable ground truth.

Observed single-cell counts and learned latent states carry measurement and model uncertainty, motivating probabilistic rather than deterministic interpretation [32]. Confounding can persist when baseline states correlate with sample identity, disease severity, exposure, or technical variables. Chemical-space limitation creates a separate uncertainty channel: a model may interpolate accurately among familiar compounds while failing on structurally distinct natural products. Model dependence should therefore be examined through sensitivity analyses, calibration where appropriate, and agreement or disagreement across reasonable representations rather than hidden behind one embedding.

Finally, uncertainty should be attached to the level at which it originates. Technical uncertainty concerns measurement and preprocessing; biological uncertainty concerns state composition and variability; model uncertainty concerns representation and prediction; interpretive uncertainty concerns whether an observed signature warrants a pathway, resistance, or translational label. Combining these into one generic confidence score would erase the very distinctions the framework is intended to preserve.

Evaluation is additionally sensitive to systematic variation, comparator choice, and the generalization scenario [26, 27, 29]. These limitations interact. Batch structure can alter state definitions; unstable states can alter inferred subpopulation selectivity; sparse time sampling can turn adaptation into apparent resistance; and pathway scores can become misleading when composition changes. None is solved merely by increasing model complexity. The framework is most defensible when uncertainty is carried forward to the response interpretation and when claims are

narrowed to the context actually tested. In vitro single-cell response remains distinct from target engagement, organism-level exposure, safety, and therapeutic effectiveness, so translational interpretation requires evidence beyond the response map itself.

CONCLUSION

Population averages remain useful summaries, but they can be pharmacologically incomplete when different cellular configurations produce similar aggregate effects. The proposed Single-Cell Response-Map Framework treats heterogeneous response as structured information: chemical context and baseline state condition perturbation, induced change is resolved across response distributions and subpopulations, pathway and trajectory layers organize interpretation, and breadth versus selectivity describes how response is distributed without assigning therapeutic value. Its contribution is therefore not a claim that single-cell modeling is universally superior, but a framework for preserving distinctions that population averaging can erase.

The framework remains computational and prospective. Its most consequential relationships require testing across chemical domains, donors, models, disease states, assays, and perturbation conditions, with uncertainty retained and mechanistic claims reserved for orthogonal experimental evidence. If those requirements are met, the average cell may become less informative than the structure of response around it.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] 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
- [2] 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
- [3] 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
- [4] 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
- [5] Peidli S, Green TD, Shen C, Gross T, Min J, Garda S, et al. scPerturb: harmonized single-cell perturbation data. *Nat Methods.* 2024;21(3):531-40. doi:10.1038/s41592-023-02144-y
- [6] Song B, Liu D, Dai W, McMyn NF, Wang Q, Yang D, et al. Decoding heterogeneous single-cell perturbation responses. *Nat Cell Biol.* 2025;27(3):493-504. doi:10.1038/s41556-025-01626-9
- [7] Burkhardt DB, Stanley JS 3rd, Tong A, Perdigoto AL, Gigante SA, Herold KC, et al. Quantifying the effect of experimental perturbations at single-cell resolution. *Nat Biotechnol.* 2021;39(5):619-29. doi:10.1038/s41587-020-00803-5
- [8] Piran Z, Cohen N, Hoshen Y, Nitzan M. Disentanglement of single-cell data with biolord. *Nat Biotechnol.* 2024;42(11):1678-83. doi:10.1038/s41587-023-02079-x
- [9] 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
- [10] Dong M, Wang B, Wei J, Fonseca AHO, Perry CJ, Frey A, et al. Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nat Methods.* 2023;20(11):1769-79. doi:10.1038/s41592-023-02040-5
- [11] Li C, Shao X, Zhang S, Wang Y, Jin K, Yang P, et al. scRank infers drug-responsive cell types from untreated scRNA-seq data using a target-perturbed gene regulatory network. *Cell Rep Med.* 2024;5(6):101568. doi:10.1016/j.xcrm.2024.101568
- [12] 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
- [13] Jiang L, Dalgarno C, Papalexi E, Mascio I, Wessels HH, Yun H, et al. Systematic reconstruction of molecular pathway signatures using scalable single-cell perturbation screens. *Nat Cell Biol.* 2025;27(3):505-17. doi:10.1038/s41556-025-01622-z
- [14] Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol.* 2022;40(2):245-53. doi:10.1038/s41587-021-01033-z

- [15] 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
- [16] 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
- [17] 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
- [18] Alsulami R, Lehmann R, Khan SA, Lagani V, Maillo A, Gomez-Cabrero D, et al. Predicting and interpreting cell-type-specific drug responses in the small-data regime using inductive priors. *Nat Mach Intell.* 2026;8(3):461-73. doi:10.1038/s42256-026-01202-2
- [19] Heumos L, Ji Y, May L, Green TD, Peidli S, Zhang X, et al. Pertpy: an end-to-end framework for perturbation analysis. *Nat Methods.* 2026;23(2):350-9. doi:10.1038/s41592-025-02909-7
- [20] Guo Y, Zhang H, Hu H, Wu J, Cao J, Hsieh CY, et al. Modelling drug-induced cellular perturbation responses with a biologically informed dual-branch transformer. *Nat Mach Intell.* 2026;8(1):96-112. doi:10.1038/s42256-025-01165-w
- [21] Driessen A, Rajwade DA, Harsanyi B, Rapsomaniki M, Born J. Conditional Monge Gap enables generalizable single-cell perturbation modelling. *Nat Mach Intell.* 2026;8(6):984-96. doi:10.1038/s42256-026-01242-8
- [22] Yu H, Qian W, Song Y, Welch JD. PerturbNet predicts single-cell responses to unseen chemical and genetic perturbations. *Mol Syst Biol.* 2025;21(8):960-82. doi:10.1038/s44320-025-00131-3
- [23] Tong X, Qu N, Kong X, Ni S, Zhou J, Wang K, et al. Deep representation learning of chemical-induced transcriptional profile for phenotype-based drug discovery. *Nat Commun.* 2024;15(1):5378. doi:10.1038/s41467-024-49620-3
- [24] Sumanaweera D, Suo C, Cujba AM, Muraro D, Dann E, Polanski K, et al. Gene-level alignment of single-cell trajectories. *Nat Methods.* 2025;22(1):68-81. doi:10.1038/s41592-024-02378-4
- [25] Wang Y, Liu X, Fan Y, Xie B, Cheng J, Wong KC, et al. Predicting drug responses of unseen cell types through transfer learning with foundation models. *Nat Comput Sci.* 2026;6(1):39-52. doi:10.1038/s43588-025-00887-6
- [26] Viñas Torné R, Wiatrak M, Piran Z, Fan S, Jiang L, Teichmann SA, et al. Systema: a framework for evaluating genetic perturbation response prediction beyond systematic variation. *Nat Biotechnol.* 2026;44(6):1050-9. doi:10.1038/s41587-025-02777-8
- [27] 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
- [28] Boyeau P, Hong J, Gayoso A, Kim M, McFaline-Figueroa JL, Jordan MI, et al. Deep generative modeling of sample-level heterogeneity in single-cell genomics. *Nat Methods.* 2025;22(11):2264-74. doi:10.1038/s41592-025-02808-x
- [29] Wei Z, Wang Y, Gao Y, Wang S, Li P, Si D, et al. Benchmarking algorithms for generalizable single-cell perturbation response prediction. *Nat Methods.* 2026;23(2):451-64. doi:10.1038/s41592-025-02980-0
- [30] Replogle JM, Saunders RA, Pogson AN, Hussmann JA, Lenail A, Guna A, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell.* 2022;185(14):2559-75.e28. doi:10.1016/j.cell.2022.05.013
- [31] Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods.* 2022;19(1):41-50. doi:10.1038/s41592-021-01336-8
- [32] Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods.* 2018;15(12):1053-8. doi:10.1038/s41592-018-0229-2