



High-Content Phenotypic Imaging as a Computational Discovery Space for Linking Natural Product Chemistry to Cellular Morphology and Mechanism of Action

João Silva^{1*}, Pedro Costa¹, Ana Beatriz²

¹Department of High-Content Phenotypic Screening, Faculty of Pharmacy, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

²Department of Cellular Morphology and Mechanism Inference, Faculty of Pharmaceutical Sciences, University of Campinas, Campinas, Brazil.

ABSTRACT

Natural products occupy chemically diverse and often sparsely annotated regions of bioactive chemical space, where a single target assay can capture only a narrow fraction of their cellular effects. High-content phenotypic imaging offers a complementary strategy by representing drug-induced cellular responses as multidimensional morphological signatures that can be searched, compared and modeled computationally. This conceptual framework examines how natural-product chemistry can be connected to image-derived cellular morphology without treating phenotypic similarity as direct evidence of molecular mechanism. It distinguishes the measured image, the computational representation derived from it, the morphology–chemistry relationship inferred by models, and the orthogonal evidence required for mechanism-of-action attribution. The framework proposes a context-indexed morphological discovery space in which compound identity, dose, exposure time, cell state, imaging protocol and representation method jointly determine the interpretable phenotype. It further positions reference compounds, chemical structures, transcriptomic or proteomic measurements and target-focused experiments as complementary evidence layers rather than interchangeable validators. Morphological similarity can therefore prioritize mechanistic hypotheses, identify activity in chemically underannotated space and guide which natural products merit deeper testing, but it can also arise from cytotoxicity, stress, convergent downstream biology or technical confounding. The proposed architecture is intended as a discovery logic rather than a validated inference rule. Its utility depends on reproducible profiling, domain-aware modeling, explicit uncertainty and orthogonal confirmation before target-level or translational conclusions are drawn.

Key Words: Cell Painting, Natural products, Morphological profiling, Phenotypic screening, Multimodal learning, Mechanism of action

eIJPPR 2025; 15(1):67-76

HOW TO CITE THIS ARTICLE: Silva J, Costa P, Beatriz A. High-Content Phenotypic Imaging as a Computational Discovery Space for Linking Natural Product Chemistry to Cellular Morphology and Mechanism of Action. Int J Pharm Phytopharmacol Res. 2025;15(1):67-76. <https://doi.org/10.51847/iBKm6KDcw1>

INTRODUCTION

Target-centric assays ask a deliberately narrow question: whether a perturbation changes a predefined molecular or functional endpoint. Phenotypic discovery starts from a different premise. A compound can produce a biologically informative cellular state before its responsible target,

pathway or polypharmacology is fully resolved, making phenotype-first evidence useful when molecular attribution is incomplete [1]. High-content imaging extends that premise because the readout is not one scalar endpoint. A single field can encode coordinated changes in organelle organization, cell shape, texture, intensity, spatial relationships and population

Corresponding author: João Silva

Address: Department of High-Content Phenotypic Screening, Faculty of Pharmacy, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

E-mail: ✉ oao.silva@ufrj.br

Received: 16 November 2024; **Revised:** 12 January 2025; **Accepted:** 17 January 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.



heterogeneity. The information content is broader, but that breadth should not be confused with mechanistic specificity.

Cell Painting operationalizes this idea by using multiplexed fluorescence imaging to generate high-dimensional profiles that support perturbation comparison, bioactivity discovery and mechanism hypothesis generation [2]. Those profiles may be constructed from engineered measurements or from learned image representations, including deep models that capture patterns not specified a priori [3]. The resulting object is therefore not simply an image and not simply a feature table. It is a computational representation of a measured cellular response whose usefulness depends on how faithfully it preserves reproducible biology while suppressing irrelevant variation.

This distinction is particularly consequential for natural products. Natural-product discovery spans structurally and biosynthetically diverse chemistry while facing persistent challenges in annotation, prioritization, dereplication and mechanistic characterization; computational methods are increasingly being used to organize and interrogate that complexity [4, 5]. The opportunity considered here is not that imaging makes natural products uniquely tractable or that it replaces target-centric pharmacology. Rather, high-content morphology can provide a searchable intermediate phenotype when chemical novelty outruns existing target annotation.

This article therefore proposes that phenotypic images be treated as a computational discovery substrate linking three separable questions: what chemical entity was tested, what cellular state was measured, and what mechanistic interpretation is justified. Keeping these questions distinct prevents a common inferential collapse. Image richness can increase the number of biologically useful comparisons available to discovery, but a morphology remains a downstream response shaped by on-target activity, off-target effects, cell context, stress, toxicity and technical conditions. The central task is to exploit that richness without assigning more mechanistic certainty than the evidence carries.

From phenotypic screening to morphological state spaces

A morphological state space can be defined as the feature or embedding space generated by a specified imaging

experiment and representation pipeline. Its geometry is therefore experimental, not intrinsic to the compound. Cell Painting optimization studies show that staining, controls, imaging, replication and processing choices materially affect profile quality [6]. In the framework proposed here, a compound occupies a morphological location only relative to a defined cell system, perturbation regime and analysis procedure. Changing those conditions can change both the apparent position of a perturbation and the meaning of its nearest neighbors.

The practical value of such spaces is demonstrated by large perturbational resources. Tens of thousands of small-molecule treatments have been converted into image-derived morphological profiles suitable for computational comparison [7]. This scale changes the role of microscopy: instead of serving only as visual confirmation, images become coordinates in a searchable perturbational atlas. The atlas can support retrieval, clustering and prioritization, but its distances are empirical relationships rather than direct measures of molecular causality.

The same logic extends beyond compound-compound comparison. Matched chemical and genetic perturbations can be profiled within a common morphological resource, creating reference relationships that support hypothesis generation across perturbation types [8]. A chemical phenotype resembling a genetic perturbation can therefore be mechanistically suggestive, yet the similarity does not by itself establish direct target engagement. Convergent pathways, compensatory responses and differences in perturbation strength can generate similar downstream states.

Recent multi-site morphology resources further show that image-derived profiles can support prediction of selected compound properties while remaining dependent on the cell line and data-generating context [9]. A useful morphological state space should therefore be understood as context-indexed: assay design determines what is visible, representation determines how it is encoded, and biological context determines which perturbations separate meaningfully.

Table 1 distinguishes the operational constructs that make a morphological state space searchable without making it mechanistically self-interpreting.

Table 1. Operational constructs of a context-indexed morphological discovery space

Construct	Meaning	Interpretive boundary
Phenotypic image	Multiplexed cellular-response measurement	Richness does not identify a unique cause
Morphological profile	Engineered or learned image representation	May encode nuisance variation
Morphological state space	Proposed context-indexed organization of profiles	Geometry depends on assay and processing
Perturbational neighborhood	Search relation among nearby profiles	Proximity does not establish target identity
Cross-perturbation reference	Chemical/genetic profile comparison	Convergent biology can create similarity

Context index	Proposed cell-system and data-generation qualifiers	Transfer requires demonstration
---------------	---	---------------------------------

Note: "Proposed" identifies article-derived synthesis rather than a validated rule.

Figure 1 visualizes why the state space is computationally useful while remaining conditional on how the phenotype was produced and represented.

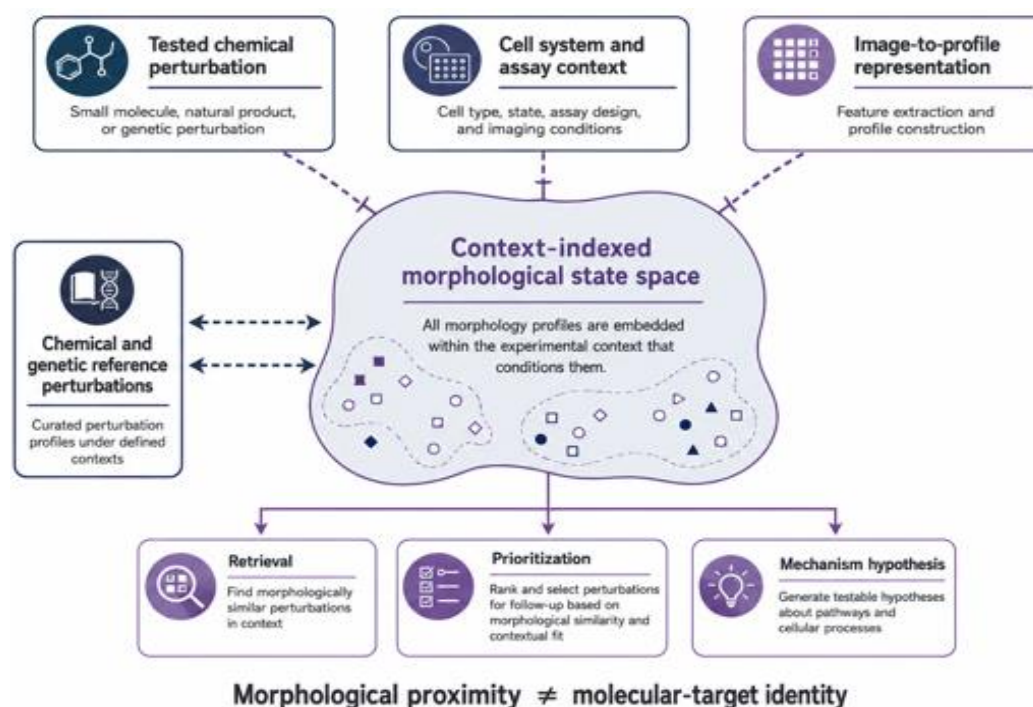


Figure 1. Morphological state space as a context-indexed computational substrate

What natural products reveal that target-centric screens can miss

Natural products are a useful stress test for morphology-centered discovery because biological activity may be evident before a single explanatory target is available. High-resolution phenotypic profiling has shown that natural products can produce distinguishable effects at the single-cell level [10]. For discovery, the important implication is not that each signature is unique or mechanistically specific. It is that cellular response can be measured and compared even when target annotations are sparse, providing an empirical layer between chemical identity and downstream mechanistic work.

Integrated activity profiling reinforces this possibility. Combining multiple functional readouts can annotate and prioritize natural products beyond what is visible in a single endpoint [11]. High-content morphology can play a similar integrative role by recording many coordinated cellular features in one assay. Yet functional annotation remains distinct from direct molecular attribution: a coherent cellular signature may arise from several targets, pathway convergence, metabolite effects or nonspecific injury.

Morphological profiling has also contributed to mechanism-oriented investigation of structurally novel pseudo-natural products when coupled to orthogonal chemical-biology experiments [12]. That sequence is important. Morphology can nominate a mechanistic neighborhood or reveal an unexpected response pattern, but the stronger inference comes from what follows—biochemical, genetic or other experiments that distinguish competing explanations. The image is therefore best treated as a hypothesis-generating layer rather than a substitute for target validation.

Natural-product activity can additionally be organized into computational maps that associate compounds with measured biological activities and facilitate prioritization [13]. Morphological state spaces extend this mapping idea from annotated endpoints to richer cellular responses. For purified natural products, this creates a route to prioritize chemically unusual compounds whose phenotypes are reproducible even when canonical target annotations are absent. For extracts, fractions or incompletely characterized samples, however, chemical attribution becomes a separate source of uncertainty; a phenotype cannot be assigned confidently to a constituent that has not been established analytically.

Building computable cellular signatures from high-content images

A phenotypic image becomes computationally useful only after it is converted into a reproducible signature. Early deep-learning work demonstrated that convolutional models could extract assay-predictive information directly from high-throughput microscopy images [14]. This establishes that image pixels can contain signal relevant to biological assay outcomes, but prediction alone does not reveal which image features are mechanistically meaningful or whether the model is exploiting nuisance correlations.

More recent representation-learning studies have shifted attention from task-specific classifiers toward embeddings designed for perturbational profiling. Learned representations can improve profile quality and perturbation retrieval on selected benchmarks [15]. For a natural-product discovery space, this matters because the desired representation is not simply the one that best separates training labels. It should retain biologically reproducible variation while remaining sufficiently robust to plate, batch, imaging and cell-state effects that would otherwise distort neighborhood structure.

Self-supervised learning further reduces dependence on predefined mechanism labels by learning useful representations from the images themselves [16]. This is attractive when natural-product annotations are sparse, but label efficiency should not be equated with biological disentanglement. A self-supervised model can still encode cell density, staining intensity, acquisition artifacts or generic stress if these dominate visual variation.

Mechanism-of-action classification from Cell Painting images shows that learned profiles can discriminate known annotated classes under benchmark conditions [17]. Such results justify using embeddings to rank mechanism hypotheses, not to declare target identity for an uncharacterized compound. Unsupervised deep phenotyping similarly demonstrates that image data can organize perturbations into biologically structured groups without complete labels [18]. These clusters become scientifically useful only when their stability, biological content and alternative explanations are evaluated.

The framework therefore treats engineered features, supervised embeddings, self-supervised representations and unsupervised latent spaces as alternative encodings of the same measured phenotype. Their value should be judged by replicate consistency, biologically meaningful retrieval, downstream task performance and resistance to nuisance variation. No representation is assumed to be intrinsically mechanistic.

Connecting chemical structure to morphological response

Chemical structure and cellular morphology are related but non-equivalent representations. Comparative modeling has shown that structural and phenotypic profiles contain partly overlapping and partly complementary information for predicting selected biological activities [19]. This complementarity is central to natural-product prioritization: structure describes molecular constitution, whereas morphology records a downstream cellular consequence in a particular context. Neither modality should be treated as a deterministic translation of the other. The practical value of the phenotypic modality is illustrated by screening studies in which Cell Painting-based models enriched active compounds for selected assays while retaining chemical diversity [20]. Such enrichment supports the use of morphology for deciding which compounds to test next. It does not imply that a predicted active has the expected mechanism, nor that morphology can replace confirmatory assays. For natural products, the same distinction is critical because structurally unusual chemistry may sit outside the training distribution even when its phenotype resembles known bioactive states.

Cross-modal learning makes the relationship more explicit. Contrastive methods can align chemical structures and bioimages in a shared representation, allowing a structure to query an image database or an image-derived state to retrieve associated chemistry [21]. The shared space is computationally powerful because it permits bidirectional search, but proximity remains statistical association. Scaffold similarity, assay composition and dataset regularities can contribute to alignment without representing causal pharmacology.

Morphological features have also been used to condition generative models that propose molecules associated with desired phenotypic information [22]. This closes a computational loop from phenotype back toward chemistry, but it remains a design hypothesis until generated molecules are synthesized and shown experimentally to produce the intended response. The proposed framework therefore treats structure-to-morphology prediction, cross-modal retrieval and morphology-conditioned generation as distinct operations. Together they make phenotype searchable from chemistry and chemistry searchable from phenotype, while preserving the requirement that experimental testing adjudicate whether the inferred relationship transfers to new natural-product chemical space.

When similar morphology suggests shared mechanism—and when it does not

Morphological similarity is most useful when interpreted as graded evidence rather than as a target label. Reference-derived subprofiles can associate an uncharacterized

perturbation with previously annotated bioactivities and thereby narrow the mechanistic search space [23]. For a natural product, a close match can therefore justify asking whether the compound perturbs a shared pathway, organelle process or functional program. It does not yet justify naming the same molecular target.

The distinction is empirical, not merely philosophical. Morphological profiling has identified common cellular modes of action among small molecules that act through different molecular targets [24]. Downstream convergence can make two perturbations look similar because both alter the same cellular process, even when their initiating interactions differ. Conversely, compounds acting on the same nominal target can diverge phenotypically because of potency, residence time, off-target activity, metabolism, cell state or exposure conditions.

Image profiling can also reveal reproducible cellular activity among compounds previously lacking conventional bioactivity annotation [25]. Such “dark” activity is valuable precisely because it identifies where additional investigation is warranted. The proposed interpretation rule is therefore asymmetric: strong phenotypic similarity can increase the plausibility of a shared biological process, whereas target identity requires evidence outside morphology. Weak similarity is likewise inconclusive because an assay may fail to visualize a relevant mechanism.

Reference compounds as mechanistic anchors

Reference compounds convert an otherwise open-ended morphological space into one containing biologically

interpretable landmarks. Mechanistically characterized compounds associated with defined cellular-injury processes can help distinguish informative perturbations from broad injury signatures and calibrate how particular forms of damage appear in high-content morphology [26]. Their role is comparative: they constrain interpretation by showing what known perturbations look like under matched conditions.

The framework proposed here treats reference compounds as anchors rather than a closed mechanistic dictionary. A close reference-derived subprofile can support annotation [23], but the coverage of any panel is necessarily limited by the compounds, mechanisms, concentrations, cell systems and time points represented. An anchor can also be polypharmacological, and its morphology may contain both mechanism-relevant and generic responses. Cellular-injury references are especially useful for identifying when an apparently distinctive phenotype resembles a broad stress state [26].

Accordingly, reference matching should answer three separate questions: whether a new profile is reproducible, whether it resembles a biologically characterized reference under comparable conditions, and whether that resemblance survives tests designed to distinguish shared downstream response from shared molecular cause. The nearest reference should therefore generate a hypothesis with an explicit coverage boundary, not terminate mechanism discovery.

Table 2 formalizes the mechanistic distinctions that become necessary once reference matching is used for natural-product interpretation.

Table 2. Interpreting reference-proximal and reference-distant morphological phenotypes

Observation	Defensible interpretation	Unresolved alternative	Next evidential step
Close annotated subprofile	Candidate shared biological process	Different upstream targets can converge	Test pathway- and target-specific predictions
Close injury reference	Injury or stress may dominate the profile	Intended pharmacology may itself induce stress	Separate dose/time injury from specific response
Similar phenotype from different targets	Downstream convergence is plausible	Hidden common off-target or context effect	Use orthogonal perturbations and target evidence
No close reference	Potentially novel or poorly covered phenotype	Weak signal, assay blindness or domain shift	Replicate, expand context and add complementary evidence

Note: The interpretation and next-step columns are proposed synthesis; they are not validated decision thresholds.

Figure 2 shows the resulting evidence gradient: reference proximity can sharpen a mechanism hypothesis, but it cannot by itself cross the boundary into target attribution.

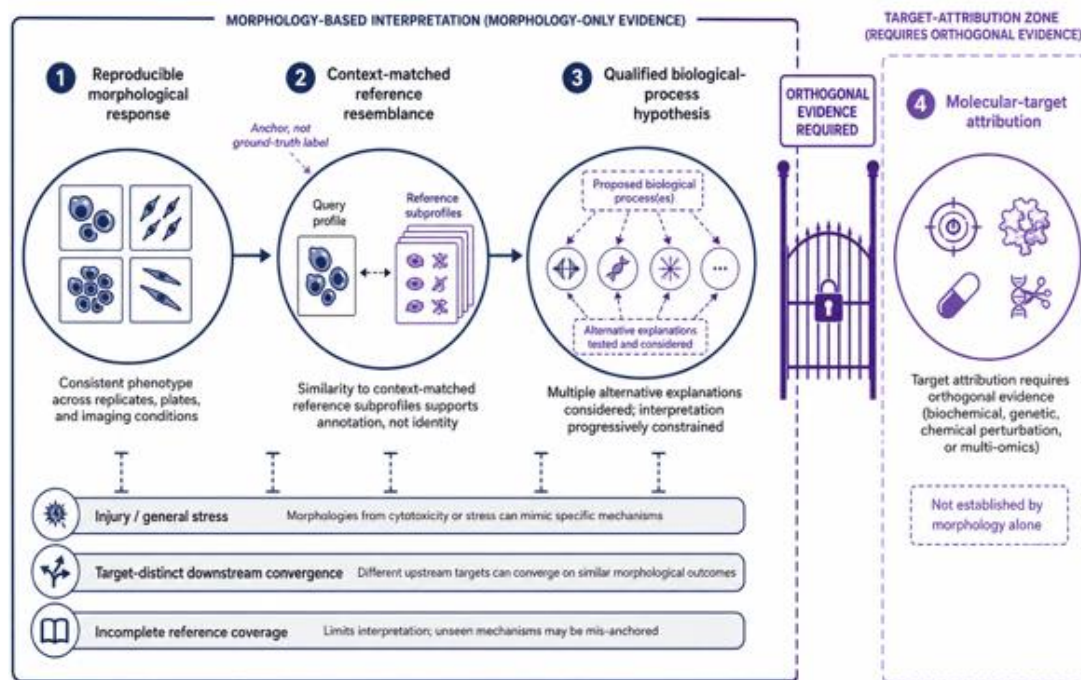


Figure 2. From morphological resemblance to qualified mechanism attribution

A chemistry–morphology–mechanism discovery architecture

The central proposal of this article is a chemistry–morphology–mechanism architecture that keeps measurement, representation and inference as distinct layers. Chemistry establishes what material was actually tested and how it is represented computationally. Morphology records a context-specific cellular response. Reference perturbations and complementary molecular data then qualify what that response may mean. Integration of morphology with chemical structure and gene expression has already shown value for selected toxicity-prediction tasks [27], providing an empirical basis for linking these modalities without treating any one as definitive.

Morphological and gene-expression profiles contain overlapping but nonredundant information, and matched perturbational resources permit their joint analysis with compound-centered representations [27–29]. The proposed architecture therefore favors triangulation rather than forced consensus. Agreement between chemical priors, morphology, transcriptomics and reference perturbations can strengthen a mechanistic hypothesis when the measurements are context-matched. Disagreement should not automatically be discarded: it may identify differences in response timing, pathway level, assay sensitivity or polypharmacology.

The architecture is deliberately bidirectional. Structure can be used to predict or retrieve phenotypic neighborhoods, while an observed phenotype can retrieve chemically distinct perturbations. Because structure and morphology

provide complementary information for selected activity predictions [19], disagreement between them is not automatically a model failure. A structurally novel natural product with a well-reproduced reference-like phenotype may be especially informative; conversely, a chemically familiar compound with an unexpected phenotype can expose context dependence or unrecognized activity. The framework preserves the provenance of each evidence layer so an analyst can distinguish conclusions supported by measured morphology, chemical similarity or computational extrapolation.

For natural products, the architecture begins with analytical confidence in compound identity and proceeds to reproducible phenotype, chemical–phenotypic comparison, reference anchoring and orthogonal evidence. Its computational outputs are rankings and hypotheses, not causal verdicts. This preserves the principal benefit of high-content imaging—searching rich cellular response space—while making the evidential boundary around mechanism explicit.

Distinguishing cytotoxicity, stress responses, and mechanism-specific phenotypes

A conspicuous phenotype is not necessarily a useful mechanism signature. Morphological profiling of PROTAC responses illustrates that desired pharmacology can coexist with toxicity- or stress-related changes in the same high-content readout [30]. Natural products pose the same interpretive problem because potent cytotoxicity, membrane disruption, mitochondrial dysfunction or other

broad injury can dominate feature space and create apparently strong similarities.

The proposed framework therefore qualifies injury before assigning mechanism specificity. Reference injury compounds can indicate whether a profile occupies a known damage neighborhood [26]. Dose and time trajectories should then be examined rather than relying on one concentration or endpoint: a specific early phenotype that later collapses into generalized injury carries a different interpretation from an immediate nonspecific stress signature. Context dependence must also remain visible because the same perturbation can appear differently across cell systems [31].

This is not a rule for deleting stress-associated profiles. A genuine pharmacological mechanism may itself trigger stress or death. The aim is to prevent generic injury from being mistaken for selective mechanistic evidence while retaining biologically meaningful toxicity as part of the compound's phenotype.

Batch effects, cell-line dependence, and image-derived confounding

Cellular morphology is conditional on the biological system used to generate it. Comparative work shows that cell-line choice alters the detectability and informativeness of phenotypic responses [32]. A morphology model trained in one cell line should therefore be regarded as operating within a defined biological domain, not as a tissue-independent representation of compound action.

Cross-cell-line live-cell profiling likewise demonstrates context-dependent responses to the same small molecules [31]. Differences in target expression, pathway wiring, cell cycle, metabolic capacity and baseline morphology can all change the observable signature. For natural products, unknown permeability or intracellular handling may further amplify these differences.

Technical variation creates a separate domain problem. Batch effects can distort phenotype-space geometry, requiring correction that reduces nuisance variation while preserving meaningful biological structure [33, 34]. Correction is therefore part of scientific validation rather than a cosmetic preprocessing step. A representation that mixes batches is unreliable, but one that erases real condition-specific biology is equally problematic. External batches, sites and cell contexts should consequently be treated as deliberate transfer tests wherever the proposed use requires them.

Active learning for selecting which natural products to test next

Once morphology becomes searchable, the discovery problem shifts from ranking existing profiles to deciding which chemistry should be measured next. Cell Painting

information has been used to expand the applicability domain of structure-based models and to support iterative selection of informative compounds [35]. This provides a methodological precedent for closed-loop testing in natural-product space.

The acquisition rule proposed here would balance chemical novelty, model uncertainty, distance from existing morphological references and experimental tractability. Natural-product computational data are heterogeneous and often sparse [5], so repeatedly selecting compounds that resemble well-characterized chemistry may improve local prediction while leaving chemically unusual regions unexplored. Conversely, novelty alone can oversample compounds whose phenotypes are weak, artifactual or analytically uncertain.

No fixed weighting is proposed. Uncertainty estimates may be poorly calibrated outside the training domain, and reference-space distance may reflect batch or context rather than useful novelty. Active selection should therefore remain an experimentally audited strategy whose value is judged by whether successive measurements improve coverage, mechanistic discrimination or prioritization on held-out chemistry.

Integrating morphology with transcriptomics, proteomics, and target evidence

Morphology should gain mechanistic value through triangulation, not by being promoted to a surrogate for every molecular modality. Cross-modal learning between Cell Painting and transcriptomics can improve selected mechanism clustering and bioactivity-modeling tasks [36]. Earlier multimodal work likewise shows that morphology and gene expression capture complementary aspects of perturbational state [28].

The proposed integration principle is context matching. Agreement is most interpretable when dose, time, cell state and perturbation identity are aligned. Discordance can be informative: transcription may precede visible morphology, proteomic or post-translational changes may occur without strong transcriptional signatures, and morphology can emphasize organellar or population-level consequences. Chemical structure contributes a prior on molecular similarity, whereas target-engagement or genetic perturbation evidence addresses different causal questions.

Accordingly, multimodal evidence should not be compressed into a universal confidence score. The framework instead asks whether independent modalities support the same biological explanation, identify complementary layers, or expose a contradiction that requires experimentation. Target-specific claims remain dependent on target-relevant evidence even when several downstream modalities agree.

Validation boundaries for mechanism-of-action inference

Validation begins by asking whether the apparent compound-level phenotype is internally coherent. Single-cell analysis can reveal heterogeneous response structure that is hidden when wells are represented only by population averages [37]. Such heterogeneity may reflect meaningful subpopulations, cell cycle or differential sensitivity, but it may also reflect segmentation and technical variation; either possibility matters before a mechanism label is assigned.

Inference is also bounded by what the assay measures. Expanded multiplexing can increase the biological detail available to Cell Painting-style profiling [38], yet additional channels do not remove causal ambiguity. An unmeasured process can remain invisible, while a highly resolved downstream consequence can still have several upstream causes.

Computational prediction adds another boundary. Generative models can forecast aspects of perturbation-induced morphology in defined settings [39], but predicted phenotype is not measured phenotype. The distinction becomes critical under chemical or biological domain shift, where plausible model output can conceal unsupported extrapolation.

Validation should also follow the level of claim. A screening claim requires evidence that a phenotype is reproducible and distinguishable from controls. A pathway hypothesis requires experiments capable of separating the nominated pathway from plausible alternatives. A target claim requires evidence that distinguishes direct or functionally necessary target involvement from downstream convergence. This claim-matched approach avoids demanding target-level proof for every useful phenotype while also preventing a high-performing classifier or visually compelling cluster from acquiring mechanistic meaning through repeated reuse. Validation is therefore not one final experiment; it is the process of matching the strength and type of evidence to the specificity of the conclusion.

The proposed minimum validation boundary is therefore claim-dependent. A reproducible morphology can establish phenotypic activity. Reference and multimodal concordance can support a mechanism hypothesis. Different targets can nevertheless converge on similar morphology [24], and reference anchors remain comparators rather than causal proof [26]. When a specific molecular target is claimed, direct or otherwise target-discriminating orthogonal evidence is required; morphology-guided studies become strongest when followed by such experiments [12]. Pathway-level or polypharmacological mechanisms may require a different validation design, but they do not justify lowering the evidential distinction between association and causation.

CONCLUSION

High-content phenotypic imaging can reposition the cell image from a descriptive endpoint to a searchable computational representation of compound-induced state. For natural products, this is valuable because cellular activity can be organized and prioritized even when structural novelty exceeds existing mechanistic annotation. The central contribution proposed here is not a morphology-to-target shortcut, but an evidence-bounded architecture linking verified chemistry, context-indexed morphology, computational representation, reference anchors and orthogonal mechanistic evidence.

Its scientific value depends on maintaining the distinctions that make the framework falsifiable. Similar images may indicate shared downstream biology without shared targets; distinctive phenotypes may be driven by injury; multimodal agreement may remain correlational; and models can fail under new chemistry, cells or batches. Reference panels are incomplete, imaging assays observe only selected biology, and prediction never substitutes for measurement.

Used within those boundaries, morphology can become a practical search layer between natural-product chemistry and mechanism: one that identifies informative cellular states, directs scarce experiments and exposes where mechanistic uncertainty remains. The next advance is therefore not simply richer imaging, but better-controlled transitions from phenotype to explanation.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic drug discovery: Recent successes, lessons learned and new directions. *Nat Rev Drug Discov.* 2022;21(12):899-914. doi:10.1038/s41573-022-00472-w
- [2] Seal S, Trapotsi MA, Spjuth O, Singh S, Carreras-Puigvert J, Greene N, et al. Cell Painting: A decade of discovery and innovation in cellular imaging. *Nat Methods.* 2025;22(2):254-68. doi:10.1038/s41592-024-02528-8
- [3] Tang Q, Ratnayake R, Seabra G, Jiang Z, Fang R, Cui L, et al. Morphological profiling for drug discovery in

- the era of deep learning. *Brief Bioinform.* 2024;25(4). doi:10.1093/bib/bbae284
- [4] Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, et al. Natural products in drug discovery: Advances and opportunities. *Nat Rev Drug Discov.* 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z
- [5] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, 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
- [6] Cimini BA, Chandrasekaran SN, Kost-Alimova M, Miller L, Goodale A, Fritchman B, et al. Optimizing the Cell Painting assay for image-based profiling. *Nat Protoc.* 2023;18(7):1981-2013. doi:10.1038/s41596-023-00840-9
- [7] Bray MA, Gustafsdottir SM, Rohban MH, Singh S, Ljosa V, Sokolnicki KL, et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *Gigascience.* 2017;6(12):1-5. doi:10.1093/gigascience/giw014
- [8] Chandrasekaran SN, Cimini BA, Goodale A, Miller L, Kost-Alimova M, Jamali N, et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nat Methods.* 2024;21(6):1114-21. doi:10.1038/s41592-024-02241-6
- [9] Wolff C, Neuenschwander M, Beese CJ, Sitani D, Ramos MC, Srovnalová A, et al. Morphological profiling data resource enables prediction of chemical compound properties. *iScience.* 2025;28(5):112445. doi:10.1016/j.isci.2025.112445
- [10] Kremb S, Voolstra CR. High-resolution phenotypic profiling of natural products-induced effects on the single-cell level. *Sci Rep.* 2017;7:44472. doi:10.1038/srep44472
- [11] Hight SK, Clark TN, Kurita KL, McMillan EA, Bray W, Shaikh AF, et al. High-throughput functional annotation of natural products by integrated activity profiling. *Proc Natl Acad Sci U S A.* 2022;119(49). doi:10.1073/pnas.2208458119
- [12] Wang L, Yilmaz F, Yildirim O, Schölermann B, Bag S, Greiner L, et al. Discovery of a novel pseudo-natural product Aurora kinase inhibitor chemotype through morphological profiling. *Adv Sci (Weinh).* 2024;11(21). doi:10.1002/advs.202309202
- [13] Lee S, van Santen JA, Farzaneh N, Liu DY, Pye CR, Baumeister TUH, et al. NP Analyst: An open online platform for compound activity mapping. *ACS Cent Sci.* 2022;8(2):223-34. doi:10.1021/acscentsci.1c01108
- [14] Hofmarcher M, Rumetshofer E, Hochreiter S, Clevert DA, Klambauer G. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. *J Chem Inf Model.* 2019;59(3):1163-71. doi:10.1021/acs.jcim.8b00670
- [15] Moshkov N, Bornholdt M, Benoit S, Smith M, McQuin C, Goodman A, et al. Learning representations for image-based profiling of perturbations. *Nat Commun.* 2024;15(1):1594. doi:10.1038/s41467-024-45999-1
- [16] Kim V, Adaloglou N, Osterland M, Morelli FM, Halawa M, König T, et al. Self-supervision advances morphological profiling by unlocking powerful image representations. *Sci Rep.* 2025;15(1):4876. doi:10.1038/s41598-025-88825-4
- [17] Wong DR, Logan DJ, Hariharan S, Stanton R, Clevert DA, Kiruluta A. Deep representation learning determines drug mechanism of action from Cell Painting images. *Digit Discov.* 2023;2(5):1354-67. doi:10.1039/D3DD00060E
- [18] Janssens R, Zhang X, Kauffmann A, de Weck A, Durand EY. Fully unsupervised deep mode of action learning for phenotyping high-content cellular images. *Bioinformatics.* 2021;37(23):4548-55. doi:10.1093/bioinformatics/btab497
- [19] Moshkov N, Becker T, Yang K, Horvath P, Dancik V, Wagner BK, et al. Predicting compound activity from phenotypic profiles and chemical structures. *Nat Commun.* 2023;14(1):1967. doi:10.1038/s41467-023-37570-1
- [20] Fredin Haslum J, Lardeau CH, Karlsson J, Turkki R, Leuchowius KJ, Smith K, et al. Cell Painting-based bioactivity prediction boosts high-throughput screening hit-rates and compound diversity. *Nat Commun.* 2024;15(1):3470. doi:10.1038/s41467-024-47171-1
- [21] Sanchez-Fernandez A, Rumetshofer E, Hochreiter S, Klambauer G. CLOOME: Contrastive learning unlocks bioimaging databases for queries with chemical structures. *Nat Commun.* 2023;14(1):7339. doi:10.1038/s41467-023-42328-w
- [22] Marin Zapata PA, Méndez-Lucio O, Le T, Beese CJ, Wichard J, Rouquié D, et al. Cell morphology-guided de novo hit design by conditioning GANs on phenotypic image features. *Digit Discov.* 2023;2(1):91-102. doi:10.1039/D2DD00081D
- [23] Pahl A, Schölermann B, Lampe P, Rusch M, Dow M, Hedberg C, et al. Morphological subprofile analysis for bioactivity annotation of small molecules. *Cell Chem Biol.* 2023;30(7):839-53.e7. doi:10.1016/j.chembiol.2023.06.003
- [24] Schneidewind T, Brause A, Pahl A, Burhop A, Mejuch T, Sievers S. Morphological profiling identifies a

- common mode of action for small molecules with different targets. *Chembiochem*. 2020;21(22):3197-207. doi:10.1002/cbic.202000381
- [25] Pahl A, Liu J, Patil S, Rezaei Adariani S, Schölermann B, Warmers J, et al. Illuminating dark chemical matter using the Cell Painting assay. *J Med Chem*. 2024;67(11):8862-76. doi:10.1021/acs.jmedchem.4c00160
- [26] Dahlin JL, Hua BK, Zucconi BE, Nelson SD Jr, Singh S, Carpenter AE, et al. Reference compounds for characterizing cellular injury in high-content cellular morphology assays. *Nat Commun*. 2023;14(1):1364. doi:10.1038/s41467-023-36829-x
- [27] Seal S, Carreras-Puigvert J, Trapotsi MA, Yang H, Spjuth O, Bender A. Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. *Commun Biol*. 2022;5(1):858. doi:10.1038/s42003-022-03763-5
- [28] Way GP, Natoli T, Adeboye A, Litichevskiy L, Yang A, Lu X, et al. Morphology and gene expression profiling provide complementary information for mapping cell state. *Cell Syst*. 2022;13(11):911-23.e9. doi:10.1016/j.cels.2022.10.001
- [29] Haghighi M, Caicedo JC, Cimini BA, Carpenter AE, Singh S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nat Methods*. 2022;19(12):1550-7. doi:10.1038/s41592-022-01667-0
- [30] Trapotsi MA, Mouchet E, Williams G, Monteverde T, Juhani K, Turkki R, et al. Cell morphological profiling enables high-throughput screening for PROTeolysis TArgeting Chimera (PROTAC) phenotypic signature. *ACS Chem Biol*. 2022;17(7):1733-44. doi:10.1021/acscchembio.2c00076
- [31] Cox MJ, Jaensch S, Van de Waeter J, Cougnaud L, Seynaeve D, Benalla S, et al. Tales of 1,008 small molecules: Phenomic profiling through live-cell imaging in a panel of reporter cell lines. *Sci Rep*. 2020;10(1):13262. doi:10.1038/s41598-020-69354-8
- [32] Heinrich L, Kumbier K, Li L, Altschuler SJ, Wu LF. Selection of optimal cell lines for high-content phenotypic screening. *ACS Chem Biol*. 2023;18(4):679-85. doi:10.1021/acscchembio.2c00878
- [33] Arevalo J, Su E, Ewald JD, van Dijk R, Carpenter AE, Singh S. Evaluating batch correction methods for image-based cell profiling. *Nat Commun*. 2024;15(1):6516. doi:10.1038/s41467-024-50613-5
- [34] Yan C, Zhang Y, Feng J, Hua H, Ruan Z, Li Z, et al. Triple-effect correction for Cell Painting data with contrastive and domain-adversarial learning. *Nat Commun*. 2025;16(1):6886. doi:10.1038/s41467-025-62193-z
- [35] Herman D, Kańduła MM, Freitas LGA, van Dongen C, Le Van T, Mesens N, et al. Leveraging Cell Painting images to expand the applicability domain and actively improve deep learning quantitative structure-activity relationship models. *Chem Res Toxicol*. 2023;36(7):1028-36. doi:10.1021/acs.chemrestox.2c00404
- [36] Ha SV, Jaensch S, Kańduła MM, Herman D, Czodrowski P, Ceulemans H. Cross modality learning of Cell Painting and transcriptomics data improves mechanism of action clustering and bioactivity modelling. *Sci Rep*. 2025;15(1):23010. doi:10.1038/s41598-025-05914-0
- [37] Stossi F, Singh PK, Marini M, Safari K, Szafran AT, Rivera Tostado A, et al. SPACE: An open-source, single-cell analysis of Cell Painting data. *Nat Commun*. 2024;15(1):10170. doi:10.1038/s41467-024-54264-4
- [38] von Coburg E, Wedler M, Muino JM, Wolff C, Körber N, Dunst S, et al. Cell Painting PLUS: Expanding the multiplexing capacity of Cell Painting-based phenotypic profiling using iterative staining-elution cycles. *Nat Commun*. 2025;16(1):3857. doi:10.1038/s41467-025-58765-8
- [39] Palma A, Theis FJ, Lotfollahi M. Predicting cell morphological responses to perturbations using generative modeling. *Nat Commun*. 2025;16(1):505. doi:10.1038/s41467-024-55707-8