Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

CELL STATE, NOT CELL TYPE, SHOULD ANCHOR COMPUTATIONAL STRATEGIES FOR THERAPEUTIC TARGET DISCOVERY

Ahmed Hamdy[1]*, Mona Khalil[2], Youssef Ibrahim[3]

1. ***Department of Cell State-Driven Target Discovery, Faculty of Pharmacy, Cairo University, Cairo, Egypt.***
2. ***Department of Computational Strategies for Cell States, Faculty of Pharmacy, Alexandria University, Alexandria, Egypt.***
3. ***Department of Beyond-Cell-Type Target Discovery, Faculty of Pharmacy, Mansoura University, Mansoura, Egypt.***

ARTICLE INFO

Received:
01 November 2024
Received in revised form:
28 January 2025
Accepted:
05 February 2025
Available online:
28 February 2025

Keywords: Single-cell transcriptomics, Cell states, Therapeutic target discovery, Disease-associated transitions, Cellular plasticity, Perturbational evidence

ABSTRACT

Single-cell technologies have increased the resolution at which disease biology can be examined, yet computational target-discovery strategies still frequently treat cell type as the principal biological unit. This practice can obscure therapeutically important variation because cells assigned to the same lineage may occupy divergent activation, differentiation, stress, resistance, or disease-associated states. The unresolved problem is therefore not simply how to classify cells more accurately, but how to determine which state-dependent molecular relationships are sufficiently reproducible, mechanistically credible, and contextually bounded to support therapeutic hypotheses. This theory article proposes State-Anchored Target Discovery as a conceptual framework for organizing such evidence. The proposed construct represents a candidate target through a state–transition–context relationship: the cellular state in which the target is implicated, the disease- or treatment-associated transition it may influence, and the biological context within which the relationship is expected to hold. It further separates state markers from causal regulators, distinguishes predicted trajectories from experimentally confirmed transitions, and treats annotation uncertainty, cross-cohort replication, and perturbational evidence as integral parts of the target claim rather than downstream technical checks. The central contribution is an evidence architecture that connects single-cell heterogeneity to target prioritization and patient-stratification hypotheses without reducing scientific adequacy to a single clustering, integration, prediction, or benchmark measure. The construct remains theoretical. It cannot establish causality, druggability, safety, developability, clinical utility, or regulatory acceptability without additional experimental and pharmaceutical evidence. Its value lies in defining a more precise analytical object for therapeutic target discovery and in clarifying the validation boundaries that should govern state-centered computational claims.

To Cite This Article: Hamdy A, Khalil M, Ibrahim Y. Cell State, Not Cell Type, Should Anchor Computational Strategies for Therapeutic Target Discovery. Pharmacophore. 2025;16(1):50-60. https://doi.org/10.51847/eBSEQIBYbb

Introduction

Single-cell transcriptomic analysis has changed the granularity at which biological variation can be represented, but increased resolution does not automatically produce therapeutically meaningful knowledge. Quality control, normalization, feature selection, dimensionality reduction, clustering, and differential-analysis choices can materially alter the cellular structures and molecular associations recovered from the same data. Consequently, a computational target-discovery claim cannot be judged solely by the apparent separation of clusters, the number of differentially expressed genes, or the internal performance of a predictive model. Analytical validity begins with an explicit account of how the chosen workflow constructs the biological units on which the claim depends [1]. This requirement is especially important in pharmaceutical research, where a computationally prominent signal may still fail to represent a stable disease mechanism, an intervention-responsive process, or a tractable therapeutic opportunity.

The challenge becomes more difficult when information is integrated across patients, technologies, tissues, disease stages, or experimental conditions. Integrative single-cell analysis can align shared biological structure and enable comparative

Corresponding Author: Ahmed Hamdy; Department of Cell State-Driven Target Discovery, Faculty of Pharmacy, Cairo University, Cairo, Egypt. E-mail: ahmed.hamdy@cu.edu.eg.

reasoning across datasets, but it can also remove genuine context-specific variation or impose equivalence between populations that should remain distinct. Integration must therefore preserve biologically meaningful differences while reducing technical variation, rather than maximizing dataset mixing as an end in itself [2]. For therapeutic target discovery, the relevant question is not merely whether cells from different datasets occupy a common embedding. It is whether the integrated representation preserves the state distinctions, transition relationships, and contextual dependencies that determine where and when a target may matter.

These analytical concerns have direct pharmaceutical consequences. Single-cell evidence may support target identification, mechanism-of-action research, biomarker development, safety assessment, and translational stratification, but its usefulness depends on the intended decision, the biological context, and the level of validation supporting the claim [3]. A target expressed in one subpopulation may represent a causal regulator, a downstream marker, a transient response, or a feature of an unrelated physiological state. Similarly, a state associated with treatment response may be useful for retrospective explanation yet unsuitable for prospective patient selection. The pharmaceutical meaning of a single-cell finding therefore emerges from the relationship among state specificity, perturbational support, replication, targetability, and decision context rather than from transcriptomic resolution alone.

This article develops the proposition that cell state, rather than cell type, should anchor computational strategies for therapeutic target discovery. Cell type remains biologically important, but it is treated here as contextual information rather than the final unit of therapeutic inference. The proposed State-Anchored Target Discovery construct represents a candidate through a state–transition–context tuple: the disease-relevant cellular state in which the candidate is implicated, the transition or plastic process it may regulate, and the tissue, treatment, donor, or microenvironmental context within which the relationship is expected to hold. The construct is intended to organize evidence, uncertainty, and validation needs. It is not presented as an empirically validated algorithm, a universal ontology, a clinical decision system, or a deployment-ready pharmaceutical workflow.

Why Cell-Type Averages Conceal Therapeutically Relevant Biology

Cell-type labels compress heterogeneous cells into broad biological categories such as epithelial cells, fibroblasts, macrophages, or T cells. This compression is useful for orientation, but it becomes problematic when the average profile of the category is treated as the therapeutic phenotype of every cell assigned to it. Pan-cancer single-cell analyses have identified recurrent transcriptional programs that vary among cells sharing the same nominal lineage and that can recur across otherwise distinct tumors [4]. Such programs may reflect proliferation, stress, inflammation, differentiation, invasion, antigen presentation, or other functional conditions that are not adequately represented by the average expression profile of the annotated cell type. A target associated with one of these programs may therefore be highly relevant to a restricted disease state while appearing modest, inconsistent, or absent in the aggregate profile.

The distinction between a cell type and a cell state is not purely terminological. A cell type usually denotes a comparatively stable lineage identity supported by developmental, anatomical, or molecular characteristics. A cell state denotes a condition that may vary within that lineage according to activation, differentiation, disease exposure, metabolic environment, treatment, or time. Disease-associated changes can therefore occur in restricted cellular neighborhoods rather than across an entire annotated population. Reference-based comparison with healthy single-cell data can identify local states that are altered in disease even when the broader cell type remains present in both healthy and affected tissues [5]. This observation supports a central theoretical claim of the present article: therapeutic inference should be attached to the altered state and its context, not automatically generalized to every cell carrying the parent cell-type label.

Cell-type averaging is particularly hazardous when therapeutically consequential states are rare, emergent, or treatment dependent. A small resistant or drug-tolerant population may contribute little to the pretreatment average but become decisive under therapeutic pressure. Single-cell analyses of cancer treatment responses have shown that transcriptional changes associated with drug tolerance and combination response can be distributed heterogeneously across individual cells [6]. The resulting signal is not simply a weaker version of the average response. It may represent a qualitatively different state with distinct survival programs, compensatory pathways, or transition potential. Averaging can consequently obscure the existence of the state, dilute its molecular features, and misdirect target selection toward pathways that characterize the dominant population rather than the population most relevant to persistence or recurrence.

A state-centered strategy should nevertheless avoid the opposite error of treating every local expression pattern as a distinct therapeutic entity. States may reflect technical variation, cell-cycle position, dissociation stress, acute exposure, stochastic expression, or arbitrary analytical resolution. The proposed construct therefore requires evidence that a state is biologically coherent, reproducible across relevant donors or cohorts, distinguishable from plausible technical explanations, and connected to a disease- or treatment-relevant transition. Cell type remains essential as a lineage and tissue constraint, but it should not be allowed to erase within-type functional heterogeneity. In this formulation, the question is not whether cell type or cell state is universally superior. It is whether state resolution provides reproducible information that changes the therapeutic interpretation of the underlying biology. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why cell-type averages conceal therapeutically relevant biology within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why cell-type averages conceal therapeutically relevant biology

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Cell-type aggregation	Does the cell-type average represent the molecular condition of most cells relevant to the therapeutic question?	Cells assigned to one lineage can express recurrent but nonuniform functional programs [4].	Establishes why a type-level summary may be an inadequate therapeutic unit.	Assuming that a statistically stable average is biologically representative of every constituent cell.	A cell-type average is informative for population-level description but does not establish within-type functional equivalence.
Disease-altered state	Is the disease signal distributed across the entire cell type or localized to a restricted neighborhood or state?	Disease-associated alterations can be detected within local cellular neighborhoods relative to healthy references [5].	Shifts inference from broad lineage difference to state-specific disease alteration.	Treating any local difference as a disease mechanism without replication or contextual control.	A disease-associated state is an analytical hypothesis until biological coherence and external reproducibility are established.
Rare or low-prevalence state	Could a small state carry disproportionate therapeutic relevance despite contributing little to the average profile?	Rare or restricted states may be diluted when expression is summarized across the parent population.	Protects potentially important resistant, pathogenic, or treatment-responsive states from averaging loss.	Equating rarity with importance or prioritizing unstable clusters because they appear distinctive.	Low prevalence alone does not establish pathogenicity, causality, or therapeutic value.
Treatment-tolerant state	Does treatment reveal or induce a state that differs from the dominant pretreatment population?	Single-cell treatment analyses can resolve heterogeneous drug-tolerance and combination-response programs [6].	Connects within-type heterogeneity to resistance and therapeutic adaptation.	Interpreting an acute stress response as a durable resistance mechanism.	A treatment-associated state requires temporal, functional, and perturbational evaluation before it is classified as a resistance driver.
State continuity and resolution	Is the proposed state discrete, continuous, overlapping, or resolution dependent?	Single-cell states may occupy gradients rather than sharply separated clusters.	Prevents the framework from equating state identity with one clustering solution.	Reifying algorithmic clusters as natural biological entities.	State boundaries must remain sensitive to analytical resolution, feature choice, and biological context.
Therapeutic relevance	Does resolving the state alter target prioritization, mechanism interpretation, or stratification hypotheses?	State resolution is useful only when it adds decision-relevant information beyond the cell-type representation.	Connects descriptive heterogeneity to the article's pharmaceutical purpose.	Assuming that higher cellular resolution necessarily improves therapeutic decisions.	State-centered analysis must demonstrate reproducible interpretive value; resolution alone is not evidence of usefulness.

Cell States, Trajectories, Plasticity, and Disease-Associated Transitions

A cell state becomes therapeutically informative when it is located within a biological process rather than treated as an isolated cluster. Disease progression, treatment exposure, differentiation, activation, adaptation, and recovery can move cells through partially ordered conditions. RNA velocity introduced a means of using relationships between unspliced and spliced transcripts to infer the local direction of short-term cellular change [7]. This approach supports a more dynamic interpretation of single-cell data by asking not only where a cell is positioned, but toward which neighboring state it may be moving. For target discovery, that directional information can help distinguish candidates associated with entry into a pathogenic state from those associated with its maintenance or exit. However, velocity remains an inference from snapshot molecular measurements and must not be equated with direct observation of lineage or time.

Trajectory inference extends this reasoning by organizing cells along putative paths, branches, or developmental progressions. Yet trajectory methods encode different assumptions about topology, starting points, branching, and continuity. Comparative evaluation has shown that method choice should be matched to the expected trajectory structure and that performance can vary substantially across biological configurations [8]. A disease-associated path inferred by one algorithm may therefore not be the unique history of the sampled cells. In the proposed framework, trajectories are treated as structured hypotheses that constrain target interpretation. They may suggest that a candidate is associated with state entry, persistence, bifurcation, or recovery, but they require corroboration through time-resolved sampling, lineage information, perturbation, or convergent biological evidence.

Plasticity further complicates the distinction between state and type because cells may enter and leave transcriptional conditions without changing lineage identity. Dynamical modeling can extend RNA-velocity analysis to transient states by estimating transcriptional kinetics and latent temporal structure [9]. This is particularly relevant to therapeutic research, where a state may emerge only during early exposure, persist under continued treatment, or disappear after withdrawal. A candidate associated with such a state could represent a regulator of adaptation, a consequence of exposure, or a temporary stress response. The state–transition–context tuple proposed here therefore includes temporal persistence and reversibility as explicit properties. A state that is stable, recurrent, and transitionally connected to disease behavior carries a different therapeutic interpretation from a short-lived molecular excursion.

Directional and dynamical models nevertheless remain conditional on data suitability, sampling density, and kinetic assumptions. Critical analyses of RNA velocity have shown that parameter non-identifiability and model misspecification can undermine direct mechanistic interpretation [10]. These limitations are not peripheral technical concerns; they determine the strength of the therapeutic claim. An inferred arrow between states does not establish that one state causes the next, that the

transition occurs in vivo, or that perturbing a correlated gene will redirect it. Accordingly, the proposed construct separates four evidentiary levels: coexistence of states, predicted directionality, experimentally supported transition, and intervention-responsive transition. Only the latter levels can begin to support claims that a candidate regulates state movement, and even then the conclusion remains context dependent and distinct from evidence of druggability, safety, or clinical value.

State-Centered Target-Discovery Principles

A state-centered target-discovery strategy begins by replacing the assumption that every therapeutically relevant population must correspond to a predefined cluster. Disease-associated changes may occupy local regions of a cellular manifold and may cut across conventional cluster boundaries. Graph-based differential-abundance methods illustrate how condition-associated changes can be localized to cellular neighborhoods without requiring all members of a broad cell type to respond uniformly [11]. The proposed State-Anchored Target Discovery construct extends this principle from population comparison to target reasoning. Its primary unit is not a gene associated with a cell type, but a **state–transition–context hypothesis** specifying the cellular condition in which the candidate is implicated, the transition it may influence, and the biological circumstances under which the relationship is expected to hold.

The second principle is that therapeutic heterogeneity should be represented directly rather than compressed into one population-wide sensitivity estimate. Computational analysis of single-cell transcriptomes can organize drug-response hypotheses around heterogeneous cellular states and identify populations predicted to differ in therapeutic susceptibility [12]. This capability is valuable because the same tumor, lesion, tissue compartment, or immune lineage may contain states with divergent response profiles. Nevertheless, predicted state-specific sensitivity is not equivalent to demonstrated target engagement or treatment efficacy. Within the proposed construct, such predictions function as hypothesis-generating evidence that must be linked to perturbation, exposure, and reproducibility before influencing target priority.

The third principle is contextual coupling. A cellular state is partly produced and maintained by interactions with neighboring cells, extracellular signals, tissue location, disease stage, and treatment history. Ligand-to-target modeling can connect signals produced by one cellular population to downstream gene programs in another, helping formulate hypotheses about how microenvironmental interactions sustain a receiver state [13]. These relationships should not be reduced to isolated ligand–receptor co-expression, nor should predicted communication be treated as observed signaling. In State-Anchored Target Discovery, extracellular context modifies the target claim: a candidate may be relevant only when a particular sender state, tissue niche, or inflammatory environment is present.

The fourth principle is intervention-aware prediction. Generative models can estimate how cellular expression states may respond to perturbation and can transfer learned responses across selected contexts [14]. Such models may help prioritize experiments, compare plausible state shifts, or identify contexts in which a perturbation is unlikely to reverse the disease-associated program. Prediction, however, remains distinct from causal confirmation. A model may reproduce an expression pattern without correctly identifying the mechanism, exposure requirement, or disease-relevant consequence. The proposed construct therefore uses predictive performance as one evidence component rather than as the final criterion of target quality. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop state-centered target-discovery principles within the article's central argument.

Table 2. Components, relationships, uncertainties, and validation needs within State-centered target-discovery principles

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
State specification	Single-cell expression or multimodal features, donor identity, disease condition, reference populations, and local cellular neighborhoods	Define the disease-relevant cellular condition without assuming that all cells carrying one type label are equivalent	A bounded and reproducible state representation	State boundaries may be continuous, overlapping, resolution dependent, or sensitive to preprocessing	Replication across analytical resolutions, donors, and orthogonal biological measurements
Differential state localization	Condition labels, neighborhood graphs, donor-aware abundance or expression models	Identify where disease- or treatment-associated changes occur within the cellular manifold	A localized state–condition association	Graph construction and sampling may alter the apparent extent of the affected state [11]	Independent-cohort replication and sensitivity analysis across reasonable graph specifications
Therapeutic heterogeneity layer	State-level expression programs, drug-response signatures, treatment context, and candidate perturbations	Represent differential therapeutic susceptibility across states	State-specific response hypotheses	Transcriptomic similarity to a drug-response signature may not reflect exposure, target engagement, or efficacy [12]	Genetic or chemical perturbation in disease-relevant models with pharmacological confirmation
Transition localization	Trajectory, velocity, temporal, lineage, or repeated-sampling information	Determine whether a candidate is associated with state entry, maintenance, branching, exit, or recovery	A state–transition relationship	Snapshot-derived paths may be non-identifiable or method dependent	Longitudinal, lineage-resolved, or time-controlled perturbational confirmation
Context-coupling layer	Cell–cell interaction models, spatial information, tissue location, treatment history, and	Specify the context in which a state–target	A context-bounded target hypothesis	Predicted intercellular communication may be	Spatial, biochemical, receptor-engagement,

	microenvironmental composition	relationship is expected to operate		indirect, incomplete, or confounded [13]	or perturbational evidence
Perturbation-prediction layer	Observed intervention responses, latent representations, cell state, dose, time, and experimental condition	Estimate plausible state-specific responses and prioritize experiments	Predicted post-perturbation state distribution	Predictive transfer can fail under unseen biology and does not establish causal mechanism [14]	External perturbation datasets, calibration analysis, and experimental target validation
Evidence-integration layer	Association, trajectory, perturbation, replication, annotation, and contextual evidence	Combine heterogeneous evidence without collapsing it into a single unqualified score	A graded state–target–context hypothesis	Evidence types are not interchangeable and may conflict	Transparent evidence profiles and prespecified rules for claim strength
Pharmaceutical boundary layer	Targetability, safety, tissue exposure, developability, assay feasibility, and clinical context	Keep biological importance separate from therapeutic feasibility	A bounded target-prioritization decision	Single-cell evidence cannot establish druggability or clinical value	Independent medicinal chemistry, pharmacology, toxicology, biomarker, and prospective studies

Separating Causal Drivers from Markers and Transient Responses

State association is insufficient for identifying a therapeutic driver. A gene may mark a disease-associated population because it is induced downstream of the underlying mechanism, because it reflects the tissue environment, or because it is expressed during a temporary stress response. Pooled CRISPR screening with single-cell transcriptomic readout provides a stronger form of evidence by connecting a defined genetic intervention to heterogeneous downstream state changes [15]. Within the proposed construct, a candidate gains causal credibility when perturbation reproducibly alters entry into, maintenance of, or exit from the relevant state. Even this evidence remains conditional on perturbation efficiency, off-target effects, compensatory mechanisms, and the biological suitability of the experimental system.

Combinatorial perturbation adds another layer because cellular states are rarely governed by one regulator acting independently. Rich single-cell phenotypes can reveal genetic-interaction manifolds in which combined interventions produce outcomes that cannot be inferred from either perturbation alone [16]. This matters for target discovery because a marker may appear central in observational data while functioning only within a redundant or compensatory network. Conversely, a modestly expressed regulator may become influential when a parallel pathway is inhibited. The proposed framework therefore distinguishes individual target association from interaction-dependent control and treats combination hypotheses as requiring direct interaction evidence rather than mere overlap between state signatures.

Genome-scale Perturb-seq demonstrates how broad genetic intervention can map genotype-to-phenotype relationships across high-dimensional cellular states [17]. Such evidence is stronger than differential expression alone because it asks whether manipulating a candidate changes the cellular phenotype rather than merely whether the candidate accompanies it. Yet scale does not remove interpretive limits. A perturbation can alter cell viability, transcriptional load, or general stress in ways that mimic disappearance of a disease-associated state. Driver classification consequently requires examination of specificity, dose or perturbation strength, rescue, temporal sequence, and convergence across independent interventions. A candidate that changes a state only through broad toxicity should not be treated as a specific regulator of that state.

Chemical perturbation is necessary because therapeutic intervention is ultimately governed by compound exposure, target engagement, dose, timing, and pharmacological context. Massively multiplexed chemical transcriptomics can resolve heterogeneous and dose-dependent responses that would be hidden in bulk measurements [18]. Such data may distinguish a durable reconfiguration of the disease-associated state from an acute exposure response, particularly when multiple time points or withdrawal conditions are examined. Nevertheless, chemical state reversal is not synonymous with therapeutic benefit. A compound may normalize a transcriptomic signature while failing to engage the intended target in tissue, achieve tolerable exposure, or improve a disease-relevant phenotype. **Table 3** organizes the evidence, constructs, and interpretation boundaries needed to develop separating causal drivers from markers and transient responses within the article's central argument.

Table 3. Components, relationships, uncertainties, and validation needs within Separating causal drivers from markers and transient responses

Evaluation dimension	What must be tested	Suitable evidence or assessment	Failure signal	Interpretive limitation	Decision relevance
State association	Whether the candidate is reproducibly associated with the specified state rather than with the parent cell type generally	Donor-aware differential analysis, neighborhood association, alternative state resolutions, and external cohorts	Association disappears after donor adjustment, reference transfer, or reasonable preprocessing changes	Association cannot distinguish cause, consequence, or correlated exposure	Supports candidate nomination only
Temporal precedence	Whether candidate activity precedes state entry or disease-relevant change	Time-course profiling, repeated sampling, lineage-resolved observation, or early perturbation readouts	Candidate expression arises only after the state or phenotype is established	Temporal order increases plausibility but does not prove causation	Helps distinguish upstream regulators from downstream markers
Genetic necessity	Whether reducing candidate activity prevents state entry,	CRISPR interference, knockout, knockdown, or other targeted loss-of-	State remains unchanged despite adequate perturbation	Redundancy or incomplete perturbation can produce false-negative results	Supports causal-driver credibility when effects are

	persistence, or phenotype expression	function experiments with single-cell readout [15]			specific and reproducible
Genetic sufficiency	Whether candidate activation induces or stabilizes the state under defined conditions	CRISPR activation, overexpression, inducible systems, and state-resolved phenotyping	Candidate activation does not induce the state or produces only nonspecific stress	Sufficiency may depend on cooperating pathways and cellular context	Identifies candidates capable of initiating or reinforcing state transitions
Interaction dependence	Whether the candidate acts independently or through compensatory and combinatorial relationships	Combinatorial perturbation and genetic-interaction mapping [16]	Apparent driver effect disappears or reverses when pathway context changes	Tested interactions represent only a subset of the relevant network	Informs combination hypotheses and resistance analysis
Phenotypic specificity	Whether perturbation changes the disease-relevant state rather than viability or transcription globally	State occupancy, transition analysis, viability controls, pathway-specific readouts, and orthogonal phenotypes	Broad cell loss or generic stress explains the apparent disappearance of the state	Transcriptomic change may remain disconnected from disease function	Prevents nonspecific toxicity from being misclassified as state control
Genome-scale causal placement	Whether the candidate's effect is distinguishable from other regulators across a broad perturbational landscape	Genome-scale Perturb-seq and comparative genotype–phenotype mapping [17]	Candidate produces no distinctive or reproducible state effect	High-dimensional perturbation does not establish druggability or safety	Helps prioritize regulators for focused mechanistic validation
Chemical convergence	Whether pharmacological intervention reproduces the genetically supported state effect	Multiple chemical probes, dose response, target-engagement evidence, and state-resolved transcriptomics [18]	Chemical and genetic interventions diverge or effects occur only at nonspecific concentrations	Compound effects may reflect polypharmacology, metabolism, or exposure artifacts	Connects biological driver evidence to a plausible intervention route
Persistence and reversibility	Whether the response is durable, reversible, or limited to acute exposure	Longitudinal profiling, washout, rechallenge, and repeated perturbation	Expression returns rapidly without changing the underlying state trajectory	Persistence in vitro may not predict persistence in tissue	Separates durable state regulation from transient responses
Rescue and orthogonal confirmation	Whether restoring candidate activity or using a distinct intervention reverses the observed effect	Rescue constructs, alternative guides, independent compounds, protein or functional assays	Effect cannot be reproduced or rescued	Rescue may be technically incomplete or pathway compensation may remain	Increases specificity and causal credibility
Context transportability	Whether the causal relationship persists across donors, models, tissues, and treatment conditions	Independent cohorts, organoids, primary cells, in vivo models, and context-interaction analysis	Effect is confined to one cell line, donor, or artificial condition	Replication supports only the tested contexts	Defines the valid scope of the target hypothesis
Pharmaceutical adequacy	Whether a causal driver is also targetable, safely modulated, and relevant at achievable exposure	Structural, chemical, safety, pharmacokinetic, pharmacodynamic, and tissue-distribution evidence	Biologically valid target lacks tractability or an acceptable therapeutic window	Causal control of a state does not establish therapeutic value	Determines whether the candidate should progress beyond biological validation

Cross-Cohort Replication, Annotation Uncertainty, and Perturbational Evidence

A state-centered target hypothesis is only as credible as the labels and boundaries used to define the state. Automated cell-identification methods differ in their dependence on reference datasets, feature representations, training labels, and the composition of the query population. Comparative analysis has shown that annotation performance varies across datasets and methods [19]. Annotation uncertainty should therefore propagate into downstream target claims. A candidate associated only with cells assigned to one label under one annotation procedure should carry lower confidence than a candidate whose state relationship persists across plausible labels, resolutions, and reference panels. Expert review can improve interpretation, but it should not conceal ambiguity or convert uncertain state membership into categorical biological certainty.

Cross-cohort analysis also depends on how technical variation is removed. Batch-correction methods differ in the balance they achieve between mixing datasets and preserving biological structure [20]. Overcorrection can erase a disease-specific or treatment-specific state, whereas undercorrection can preserve technical clusters that resemble biological populations. Consequently, replication should not be defined simply as successful co-embedding of cells from multiple cohorts. A state–target relationship should remain interpretable in unintegrated data, survive more than one reasonable integration strategy, and reproduce at the donor or cohort level. Integration is part of the evidence-generating model and must be evaluated as such.

Atlas-scale benchmarks reinforce the need for multidimensional evaluation. Effective integration requires assessment of both batch removal and biological conservation, and no single metric can establish whether the resulting representation is transportable to a new therapeutic context [21]. High dataset mixing may coexist with loss of rare states, distorted neighborhood relationships, or weakened donor-level effects. Conversely, preservation of every local difference may leave technical structure unresolved. The proposed framework therefore treats cross-cohort replication as a profile of evidence: state recovery, feature concordance, effect-direction agreement, donor-level support, sensitivity to integration choice, and preservation of the transition or perturbation relationship.

Reference mapping can provide a common coordinate system for comparing new samples with established cellular atlases and for incrementally expanding those atlases [22]. Its value is conditional on reference coverage. A query state that is absent from the reference may be incorrectly forced into the nearest known category, while a disease-specific population may be interpreted as an annotation error. Perturbational evidence should therefore be evaluated alongside reference transfer rather than after annotation has been treated as settled. **Figure 1** presents the cell-state therapeutic target atlas, showing how the article's key components and boundaries are connected within cross-cohort replication, annotation uncertainty, and perturbational evidence.

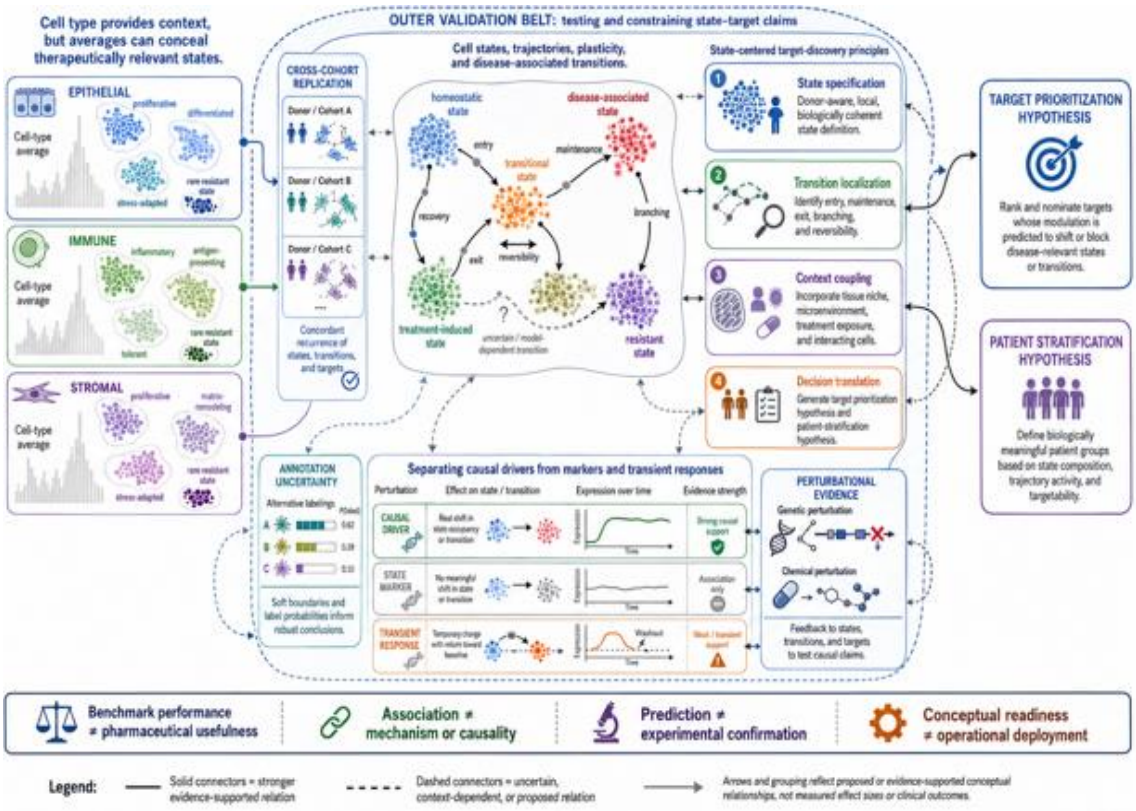

Figure 1. Cell-State Therapeutic Target Atlas.

The figure is an original conceptual synthesis that organizes the article's central contribution across cell-type averages conceal therapeutically relevant biology, cell states, trajectories, plasticity, and disease-associated transitions, cell states, disease-associated transitions, state-centered target-discovery principles, separating causal drivers from markers and transient responses. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Implications for Target Selection and Patient Stratification

Single-cell studies of cancer have demonstrated the interpretive value of resolving malignant, immune, and stromal heterogeneity while also exposing limitations related to sampling, state definition, temporal interpretation, and clinical translation [23]. These lessons apply more broadly to state-centered target discovery. A target should not be advanced merely because it is enriched in a disease-associated state. The state must be sufficiently prevalent or consequential to matter, the target must influence a relevant transition or phenotype, and the relationship must remain stable across appropriate biological contexts. Target selection should therefore remain a staged process in which state evidence informs biological prioritization but remains separate from assessments of tractability, safety, exposure, and development feasibility.

Patient stratification poses an additional aggregation problem because cell-level states must ultimately be converted into reproducible patient-level variables. Specific T-cell states have been associated with response to checkpoint immunotherapy in melanoma, supporting the possibility that state composition may inform treatment-response hypotheses [24]. Such association does not alone establish a clinically useful biomarker. A stratification strategy must define how cellular information is aggregated, how uncertain or rare states are handled, whether the state can be measured using a feasible assay, and whether the resulting patient-level classification is externally calibrated. Thresholds derived after observing outcomes should not be mistaken for prospectively validated decision rules.

The proposed construct may also support target prioritization by linking single-cell states to molecular networks and disease-level evidence. A single-cell-based strategy integrating cellular and network information has been used to prioritize diagnostic and therapeutic hypotheses in complex disease contexts [25]. That study-specific validation cannot be transferred automatically

to the present theory or to unrelated diseases. In State-Anchored Target Discovery, network evidence is used to position a candidate within the regulatory structure of a defined state, but it does not supersede perturbation. A highly connected target may be a driver, a downstream hub, or an essential physiological regulator whose intervention would be unsafe. Network placement should therefore refine experimental priorities rather than function as a universal ranking score.

Resistance-state analysis illustrates both the potential and the limitations of state-centered translation. Single-cell profiling can resolve rare or pre-existing populations associated with resistance to targeted therapy and can identify candidate markers that would be obscured in aggregate measurements [26]. Such markers may support monitoring, combination design, or mechanistic experiments, but they require separate evidence of causal control, targetability, and clinical relevance. The practical implication is a bounded decision architecture: state evidence may support prioritization of what to test, in which population, and under which treatment context, but it cannot determine by itself whether a target or biomarker should enter clinical use. **Table 4** organizes the evidence, constructs, and interpretation boundaries needed to develop implications for target selection and patient stratification within the article's central argument.

Table 4. Evaluation, implementation, and research priorities arising from Cell State, Not Cell Type, Should Anchor Computational Strategies for Therapeutic Target

Research or implementation priority	Unresolved gap	Required methodological work	Evidence needed	Responsible actors	Expected contribution
Standardize state specification	States are often defined using incompatible features, resolutions, and annotation conventions	Develop multi-resolution, donor-aware representations with explicit uncertainty and biologically interpretable anchors	Cross-platform datasets, alternative annotations, multimodal measurements, and independent replication	Bioinformaticians, single-cell biologists, disease experts, and ontology developers	More reproducible definitions of the cellular unit supporting the target claim
Preserve disease-relevant variation during integration	Harmonization may erase rare or context-specific states or preserve technical structure	Compare integrated and unintegrated analyses and evaluate biological conservation alongside batch removal	Multi-site cohorts, technical controls, known biological differences, and sensitivity analyses	Computational method developers, statisticians, and data-generating laboratories	Integration procedures aligned with the intended biological claim
Establish donor-level replication	Cell-level sample size can obscure limited biological replication	Use hierarchical, pseudobulk, neighborhood, or mixed models that preserve donor as the inferential unit	Independent donors, cohorts, disease stages, and treatment contexts	Biostatisticians, clinical investigators, and cohort consortia	Reduced pseudoreplication and more defensible state–target associations
Validate transitions longitudinally	Snapshot trajectories cannot uniquely establish state progression or reversibility	Combine trajectory inference with temporal, lineage, clonal, or repeated-sampling designs	Time-course samples, lineage tracing, treatment withdrawal, and rechallenge	Experimental biologists, translational researchers, and computational modelers	Distinction among coexistence, progression, adaptation, and recovery
Test causal control of states	Many nominated targets remain correlational markers	Design genetic and chemical perturbations around state entry, maintenance, exit, and disease-relevant phenotype	Loss- and gain-of-function studies, rescue, orthogonal interventions, dose and timing analyses	Functional genomics teams, pharmacologists, and disease-model specialists	Separation of causal state regulators from markers and transient responses
Link state biology to targetability	Biological relevance is frequently conflated with therapeutic feasibility	Add independent assessment of structure, chemistry, target engagement, tissue exposure, safety, and developability	Chemical matter, pharmacokinetic and pharmacodynamic studies, toxicology, and tissue distribution	Medicinal chemists, structural biologists, pharmacologists, toxicologists, and formulation scientists	More realistic progression from state biology to pharmaceutical target selection
Develop patient-level state variables	Cell-level associations do not directly produce stable patient classifications	Define aggregation rules, uncertainty handling, assay transfer, thresholds, and calibration procedures	External cohorts, feasible assays, prevalence estimates, comparator models, and outcome linkage	Clinical scientists, pathologists, statisticians, assay developers, and regulators	Testable patient-stratification hypotheses with explicit intended use
Evaluate population transportability	State prevalence and target relevance may vary across demographic and clinical contexts	Conduct prespecified transportability analyses across diverse populations and treatment settings	Multi-population cohorts with harmonized clinical and biological metadata	Clinical consortia, epidemiologists, community stakeholders, and statistical geneticists	Explicit boundaries for reuse and reduced risk of biased stratification
Build perturbation-reference resources	Existing references incompletely cover diseases, doses, times, and cellular contexts	Generate standardized genetic and chemical perturbation atlases with matched metadata and quality controls	Multi-dose, multi-time, multi-model perturbation experiments	Academic consortia, pharmaceutical scientists, data standards groups, and funders	Stronger comparison of predicted and observed state responses
Prospectively evaluate decision value	Retrospective associations rarely show that state-informed decisions	Predefine target-selection or stratification rules and compare them with existing approaches	Prospective studies, external calibration, intervention linkage, and decision-relevant outcomes	Translational teams, trialists, statisticians, ethics bodies, and regulatory scientists	Evidence of whether state-centered reasoning adds value beyond established practice

	improve research or clinical outcomes

Figure 2 presents the evidence, validation, and translation boundary observatory for cell state, not cell type, should anchor, showing how the article's key components and boundaries are connected within implications for target selection and patient stratification.

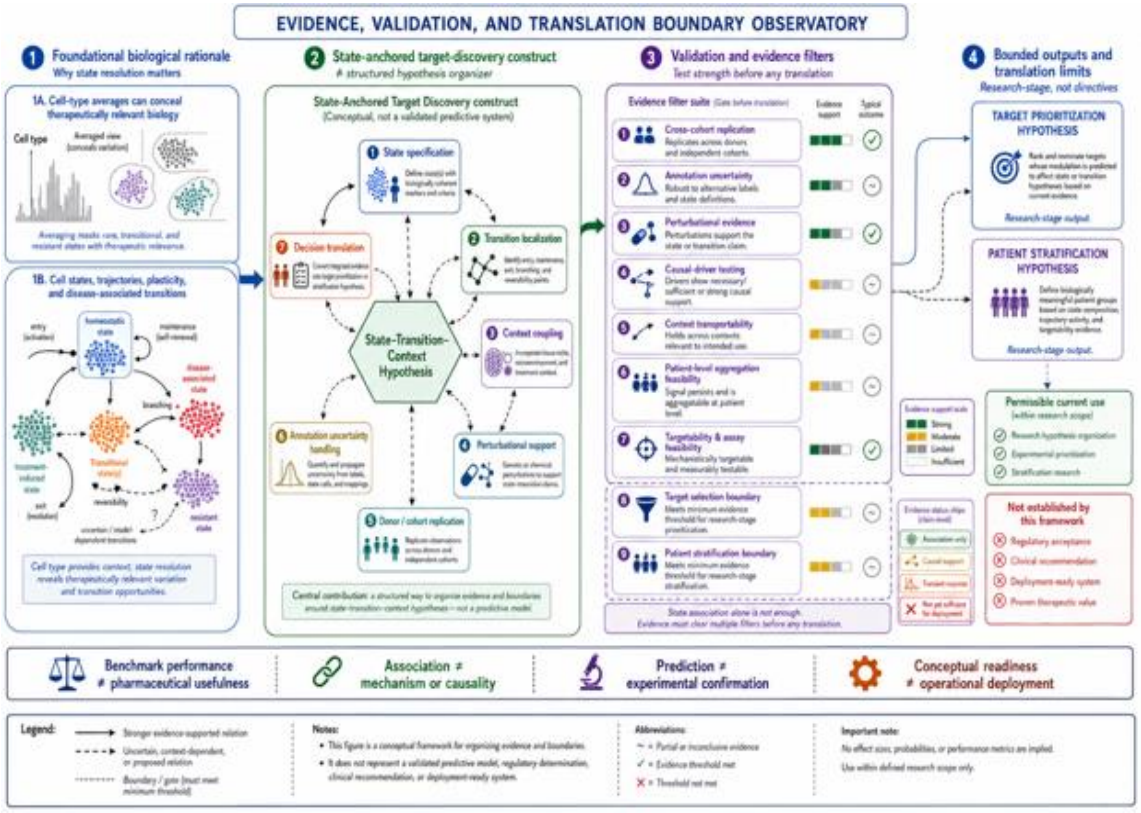

Figure 2. Evidence, Validation, and Translation Boundaries.

The figure is an original conceptual synthesis that shows how claims move from conceptual plausibility through evaluation, uncertainty assessment, and bounded pharmaceutical use. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Limitations and Boundary Conditions

The first limitation is conceptual. Cell states do not possess universally accepted boundaries, and the same biological population may be represented as a cluster, a neighborhood, a gradient, a mixture, or a transient position depending on the analytical purpose. State-Anchored Target Discovery does not resolve this ontological problem by declaring one representation correct. Instead, it requires the representation to be explicit, uncertainty aware, and appropriate to the therapeutic claim. A state that is useful for identifying a resistance transition may not be the appropriate unit for safety assessment or patient classification. The construct is therefore conditional on the scale at which the disease mechanism and intended decision operate.

The second limitation concerns evidence. Single-cell observations are vulnerable to sampling bias, dissociation effects, technical variation, incomplete tissue coverage, uncertain annotation, and limited donor replication. Trajectory and perturbation models introduce further assumptions concerning kinetics, topology, intervention specificity, dose, and experimental context. Cross-cohort recurrence may increase confidence, but it does not eliminate confounding or prove causality. Likewise, perturbational evidence obtained in a cell line or ex vivo model may not reproduce in intact tissue. The framework can organize the strength and scope of these claims, but it cannot manufacture missing temporal, mechanistic, pharmacological, or clinical evidence.

The third limitation is translational. A state-linked causal regulator may be undruggable, systemically toxic, inaccessible in the relevant tissue, pharmacologically redundant, or impractical as a biomarker. Conversely, a therapeutically useful target may not be the most state-specific molecular feature. The proposed construct must therefore remain upstream of, and connected to, independent evaluation of medicinal chemistry, target engagement, exposure, safety, developability, assay feasibility, and prospective utility. It must not be used as a clinical recommendation, regulatory classification, automatic target-ranking system, or substitute for multidisciplinary scientific judgment.

Research Agenda and Implementation Priorities

The first research priority is to make state-centered claims falsifiable. Future studies should specify how states were defined, which alternative representations were tested, how donor-level variation was handled, and what evidence would cause the state–target hypothesis to be rejected. Evaluation should compare state-aware models with appropriate cell-type-level baselines rather than assume that additional resolution is beneficial. Such comparisons should examine whether state information improves reproducibility, causal interpretation, perturbation selection, or patient-level decision support without relying on one aggregate performance measure. Negative results should be retained because failure to reproduce a state or intervention response defines the boundary of the construct.

The second priority is to create evidence infrastructure that connects observation to intervention. This requires longitudinal and lineage-resolved datasets, standardized genetic and chemical perturbation references, matched spatial and multimodal measurements, and metadata describing dose, timing, tissue context, patient characteristics, and treatment history. Reference atlases should permit unknown or out-of-reference states rather than forcing every query into an existing category. Reporting practices should disclose annotation uncertainty, integration sensitivity, donor-level support, model assumptions, and the distinction between cell-level and patient-level inference. Such infrastructure would not validate State-Anchored Target Discovery automatically, but it would make its components testable.

The third priority is staged implementation. Initial use should be limited to research hypothesis organization: defining which state, transition, context, and perturbational evidence support a candidate. Progression to target prioritization should require replicated causal evidence and an independent targetability assessment. Progression to stratification research should additionally require a feasible assay, patient-level aggregation rules, external calibration, and a predefined intended use. Prospective pharmaceutical or clinical evaluation should occur only after these earlier stages are satisfied. This staged approach protects against the premature conversion of descriptive cellular heterogeneity into operational therapeutic decisions.

Conclusion

Cell type remains indispensable for describing lineage, anatomy, and broad biological function, but it is often too coarse to serve as the principal unit of therapeutic target inference. The proposed State-Anchored Target Discovery construct instead organizes a candidate around the cellular state in which it operates, the disease- or treatment-associated transition it may influence, and the context within which that relationship is expected to hold. Its central contribution is not a new performance score or validated target-ranking algorithm, but an evidence architecture that separates state association from causal regulation, prediction from experimental confirmation, replication from forced harmonization, and biological importance from pharmaceutical usefulness. The construct may support more precise target and patient-stratification hypotheses, but it cannot establish targetability, safety, developability, clinical utility, regulatory acceptability, or deployment readiness. Its scientific value will depend on whether future studies can operationalize its components, expose them to falsifiable evaluation, and demonstrate that state-centered reasoning adds reproducible decision value beyond cell-type-based analysis.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: A tutorial. Mol Syst Biol. 2019;15(6). doi:10.15252/msb.20188746
2. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20(5):257-72. doi:10.1038/s41576-019-0093-7
3. Van de Sande B, Lee JS, Mutasa-Gottgens E, Naughton B, Bacon W, Manning J, et al. Applications of single-cell RNA sequencing in drug discovery and development. Nat Rev Drug Discov. 2023;22(6):496-520. doi:10.1038/s41573-023-00688-4
4. 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
5. Dann E, Cujba AM, Oliver AJ, Meyer KB, Teichmann SA, Marioni JC. Precise identification of cell states altered in disease using healthy single-cell references. Nat Genet. 2023;55(11):1998-2008. doi:10.1038/s41588-023-01523-7
6. Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun. 2021;12(1):1628. doi:10.1038/s41467-021-21884-z
7. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8. doi:10.1038/s41586-018-0414-6

8. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019;37(5):547-54. doi:10.1038/s41587-019-0071-9
9. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;38(12):1408-14. doi:10.1038/s41587-020-0591-3
10. Gorin G, Fang M, Chari T, Pachter L. RNA velocity unraveled. PLoS Comput Biol. 2022;18(9). doi:10.1371/journal.pcbi.1010492
11. 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
12. Fustero-Torre C, Jiménez-Santos MJ, García-Martín S, Carretero-Puche C, García-Jimeno L, Ivanchuk V, et al. Beyondcell: Targeting cancer therapeutic heterogeneity in single-cell RNA-seq data. Genome Med. 2021;13(1):187. doi:10.1186/s13073-021-01001-x
13. Browaeys R, Saelens W, Saeys Y. NicheNet: Modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159-62. doi:10.1038/s41592-019-0667-5
14. 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
15. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods. 2017;14(3):297-301. doi:10.1038/nmeth.4177
16. Norman TM, Horlbeck MA, Replogle JM, Ge AY, Xu A, Jost M, et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science. 2019;365(6455):786-93. doi:10.1126/science.aax4438
17. 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
18. Srivatsan SR, McFaline-Figueroa JL, Ramani V, Saunders L, Cao J, Packer JS, et al. Massively multiplex chemical transcriptomics at single-cell resolution. Science. 2020;367(6473):45-51. doi:10.1126/science.aax6234
19. Abdelaal T, Michielsen LCM, Cats D, Hoogduin D, Mei H, Reinders MJT, et al. A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol. 2019;20(1):194. doi:10.1186/s13059-019-1795-z
20. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. 2020;21(1):12. doi:10.1186/s13059-019-1850-9
21. 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
22. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Büttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol. 2022;40(1):121-30. doi:10.1038/s41587-021-01001-7
23. Suvà ML, Tirosh I. Single-cell RNA sequencing in cancer: Lessons learned and emerging challenges. Mol Cell. 2019;75(1):7-12. doi:10.1016/j.molcel.2019.05.003
24. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell. 2018;175(4):998-1013. doi:10.1016/j.cell.2018.10.038
25. Gawel DR, Serra-Musach J, Lilja S, Aagesen J, Arenas A, Asking B, et al. A validated single-cell-based strategy to identify diagnostic and therapeutic targets in complex diseases. Genome Med. 2019;11(1):47. doi:10.1186/s13073-019-0657-3
26. Ho YJ, Anaparthy N, Molik D, Mathew G, Aicher T, Patel A, et al. Single-cell RNA-seq analysis identifies markers of resistance to targeted BRAF inhibitors in melanoma cell populations. Genome Res. 2018;28(9):1353-63. doi:10.1101/gr.234062.117