Review Article

Machine learning to dissect perturbations in complex cellular systems

Pablo Monfort-Lanzas [a,b] , Katja Rungger [a] , Leonie Madersbacher [a] , Hubert Hackl [a,*,1,2]

[a] *Institute of Bioinformatics, Biocenter, Medical University of Innsbruck, Austria*
[b] *Institute of Medical Biochemistry, Biocenter, Medical University of Innsbruck, Austria*

A B S T R A C T

Understanding the responses of biological systems to various perturbations, such as genetic, chemical, or environmental challenges, is essential for reconstructing causal network models. Emerging single-cell technologies have become instrumental in elucidating cell states and phenotypes and they have been used in combination with genetic screening. Recent advances in machine learning and artificial intelligence architectures have stimulated the development of computational tools for modeling perturbations and the response to compounds. This study outlined core principles underpinning perturbation analysis and discussed the methodologies and analytical frameworks used to decode drug and genetic perturbation responses, complex multicellular interactions, and network dynamics. The current tools used for various applications were overviewed. These developments hold great promise for improving drug development and personalized medicine. Foundation models and perturbation cell and tissue atlases offer immense potential for advancing our understanding of cellular behavior and disease mechanisms.

1. Introduction

Complex biological systems are organized hierarchically, from the molecular level to tissues or heterogeneous cellular structures. Emerging single-cell technologies have facilitated the elucidation of cell states and phenotypes. In addition to methods analyzing the transcriptomics layer, such as single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics, technologies for the comprehensive analyses of proteins, metabolites, multiomics integration, and imaging have been developed. However, dynamic biological systems are organized through a complex network of interconnected pathways that maintain cellular homeostasis. To obtain a mechanistic understanding of system behavior, termed systems identification in technical system analysis, it is necessary to perturb the system. Understanding the responses of biological systems to various perturbations, such as genetic, chemical, or environmental challenges, is essential for reconstructing causal network models. This can be achieved through the rational manipulation of cells, which is crucial in modern biomedical research [1,2].

This development has significantly influenced precision medicine, in which chemical and therapeutic interventions are studied in disease models such as organoids. Perturbation modeling has become critical for

developing personalized treatments, controlling disease progression, and identifying therapeutic targets. It is also essential for precise engineering of cells for regenerative therapies. Experimental methods such as CRISPR or RNAi technologies permit the genetic manipulation of individual genes and enable the performance of pooled genetic screens across many genes in combination with readout from scRNA-seq (Perturb-Seq), providing ideal approaches to address the perturbation response in various applications [3].

Given the high number of potential molecular changes and their often non-linear responses, perturbation analysis possesses high combinatorial complexity. Different modeling approaches capture the responses of biological systems to perturbations. Network analyses map key interactions and regulatory patterns [4–6]. Dynamic logic models [7–9] describe system-wide changes without requiring detailed kinetic data. Meanwhile, differential equation models [6,10] provide precise, quantitative predictions about the evolution of systems over time, but they rely on extensive parameterization. However, computational advancements, particularly artificial intelligence (AI) and machine learning (ML), are crucial for overcoming these challenges. AI/ML approaches trained on high-content pooled screens are increasingly feasible and cost-effective, providing a foundation for robust

computational models [1]. AI based methods are also pivotal for analyzing perturbation proteomics data [11]. Although many tools have been developed to study the response genetic intervention, tools for assessing the responses to pharmacologic compounds and dose responses at the single-cell single-cell level are limited. Therefore, it is critical to perform perturbation modeling at the single-cell level to evaluate the effect of the specific drug concentrations such as benchmark doses as used in toxicology or to model treatment resistance. Therefore, large perturbation datasets and systematic repositories for biological responses to treatment, such as Connectivity Map [12], are indispensable sources that can be used by transfer learning or deep generative models to identify drug response at the single-cell level [13, 14].

In this review, we outline the core principles underpinning perturbation analysis and discuss the methodologies and analytical frameworks, especially AI/ML, used to decode drug and genetic perturbation responses, complex multicellular interactions, and network dynamics We also provide an overview of the current tools and software used for various applications.

2. Principles of perturbation modeling

When a perturbation, such as a genetic alteration, pharmacological treatment, or environmental stressor, is introduced, cellular homeostasis is disrupted, leading to cascades of molecular changes. Perturbations can be broadly categorized into intrinsic and extrinsic types (Fig. 1). Intrinsic perturbations, such as genetic mutations, gene deletions, and transgene insertions originate within the organism. By contrast, extrinsic perturbations arise from external influences, such as exposure to drugs, cytokine treatments or co-culture conditions replicating the tumor microenvironment [15]. The primary goal of perturbation analysis is mapping these molecular changes to understand their effects on cellular function, signaling pathways, and phenotypic outcomes. Previously, four main objectives in single-cell perturbation modeling were described [2,15]. These objectives are defined as solvable tasks, each paired with specific evaluation metrics that can be used to assess model performance and guide research in the field [15]. First, a key goal is predicting the responses of cells to various perturbations, including novel changes in gene, protein, and metabolite levels, which can help identify new cell states and markers of drug sensitivity, as evaluated using regression between predicted and observed responses. Second, understanding a compound's mode of action (MoA), including its molecular targets and underlying biological mechanisms, is crucial for identifying novel compounds or repurposing them in drug discovery. To evaluate the model predictions of affected targets and pathways, classification metrics such as precision and recall can be used. Third, modeling the interactions between genetic and chemical perturbations is essential for developing effective combination therapies. The synergy of different compounds can be estimated using categorical or continuous based multivariate modeling. Lastly, perturbation modeling also extends to predicting chemical structures and generating new small molecules with different biological effects. Similarity scores or classification metrics are used to compare predicted compounds and evaluate their structural features. Together, these objectives provide a comprehensive
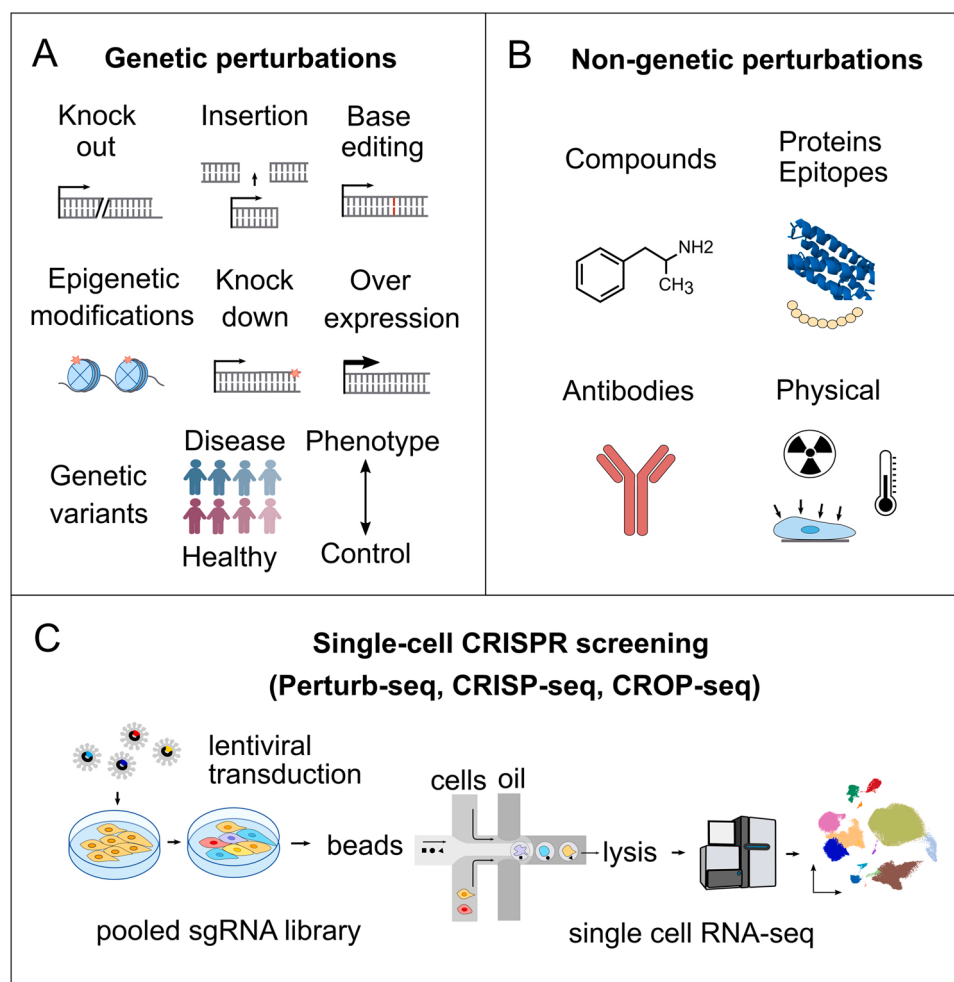

Fig. 1. Genetic (A), nongenetic perturbations (B), and methods for genetic manipulation (CRIPSR/Cas9, RNAi) and scRNA-seq read out (C).

framework for advancing single-cell perturbation studies and drug discovery.

A critical aspect of perturbation analysis is understanding the reorganization of molecular networks in response to stimuli [15]. Numerous techniques and models have been developed to analyze perturbation data, addressing the complexity of biological systems (Fig. 2). The most common methods employed in single-cell perturbation analysis are classical statistical and ML inference models, variational autoencoders (VAEs), graphical models, and transformer architectures often applied in foundation models [2,16] (Table 1).

Classical statistical and ML methods include linear regression, generalized linear models, random forests (RFs), and support vector machines. These methods have been widely used because of their interpretability, and they serve as important baselines for more complex models. For example, scDist uses a linear mixed-effects model to detect transcriptomic differences between conditions, accounting for individual and technical variability to identify cell-specific perturbation responses [17]. Augur uses an RF classifier to rank cell types based on their perturbation sensitivity, calculating the cross-validated area under the receiver operating characteristic curve to distinguish conditions [18]. MIMOSCA integrates linear regression models with interaction terms to analyze perturbation effects while controlling for covariates and minimizing confounding effects [3]. However, classical models struggle to capture non-linear relationships, cell-cell dependencies, and dynamic

cell-state transitions [19,20]. These limitations make advanced architectures such as VAEs, graphical neural networks, and transformers essential for single-cell perturbation analysis.

Gene-regulatory networks (GRNs) can be combined with graphical neural networks to build the foundation for interpretable approaches that represent gene expression regulation as networks or graphs, incorporating components such as transcription factors, splicing factors, noncoding RNAs, and metabolites [21]. GRNs are crucial for perturbation modeling because they capture gene-gene interactions and help predict changes in gene expression during perturbations [2,16]. For example, CellOracle constructs GRNs by integrating single-cell RNA-seq and ATAC-seq data to model transcriptional connections and permits the study of transcription factor-driven cell fate transitions [22]. Graph-enhanced gene activation and repression simulator uses unperturbed single-cell gene expression data and a specified perturbation set to predict transcriptional responses to gene perturbations [23]. Single cell integrative GRN inference constructs cell type-specific GRNs from scRNA-seq and spatial transcriptomics data by employing gradient boosting regression to infer regulatory relationships [24]. Importantly, these methods both predict cellular responses to perturbations and make them interpretable.

Dimensionality reduction is essential for analyzing high-dimensional single-cell data. Principal component analysis is often used for this purpose, but has limitations in correctly representing global and local
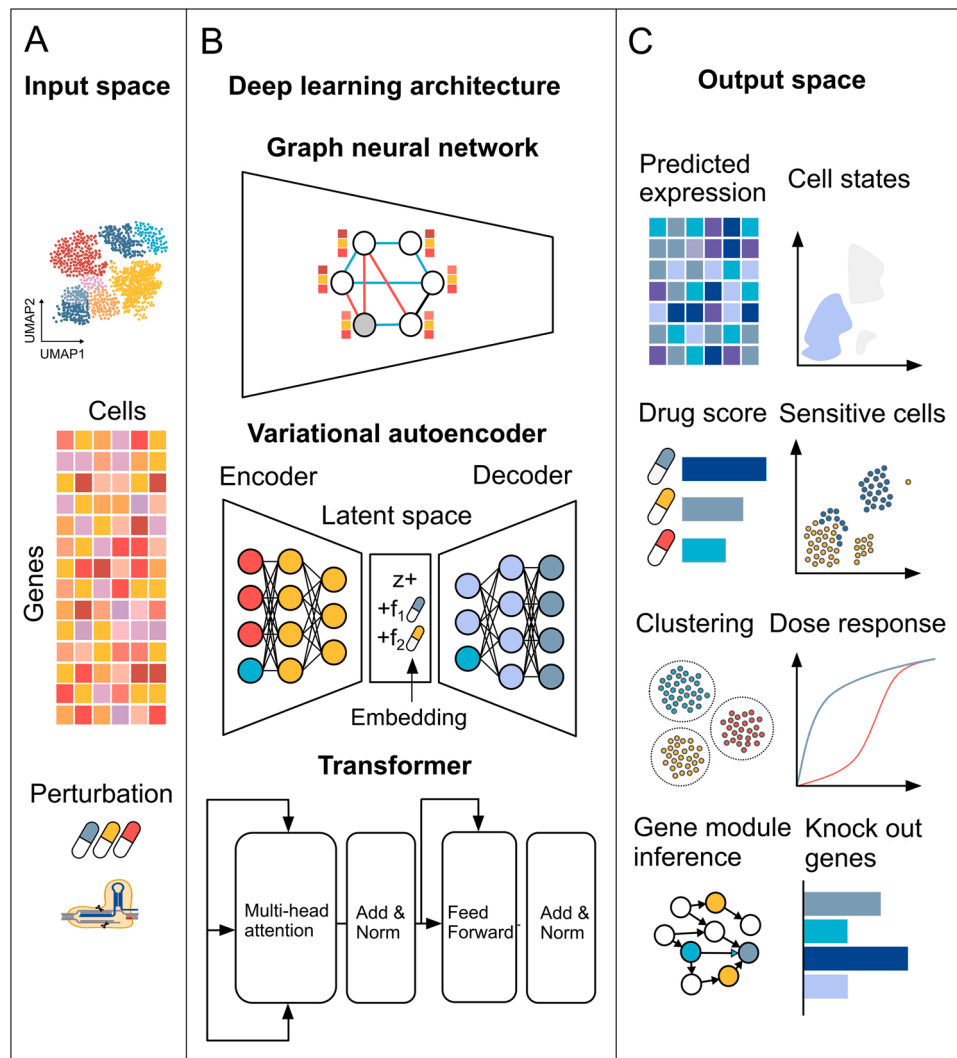

Fig. 2. Overview of ML and network models for scRNA-seq perturbation analysis. The schematic illustrates input space (A) various architectures, including graphical neural networks, variational autoencoder, and transformer (B), and output space (C).

Table 1

Computational and experimental tools for perturbation analyses.

Tool name	Ref.	Input data	Model/method	Description
Drug response				
ASGARD	[82]	scRNA-seq	Predictive pipeline	Matches diseased clusters with normal clusters to score drugs that reverse disease-related expression.
Beyondcell	[83]	scRNA-seq	Therapeutic clustering	Identifies tumor subpopulations with unique drug responses.
CaDRReS-Sc	[84]	scRNA-seq	Recommender systems	Estimates IC50s for cell clusters from gene expression data.
DREEP	[58]	scRNA-seq	Functional enrichment analysis	Predicts drug sensitivity from scRNA-seq data using pharmacogenomic screens, validated across cancer datasets
DrugCell	[85]	Tumor cell mutation and chemical data	Dual-branch deep learning	Models of mutation impact on pathways to predict drug sensitivity.
PRnet	[13]	Bulk and scRNA-seq	Perturbation-conditioned generative model	Integrates bulk and single-cell data via harmonization and transfer learning to predict single-cell drug responses and resistance mechanisms.
scDEAL	[14]	Bulk and scRNA-seq	Deep transfer learning	Uses bulk data to predict single-cell drug response, highlighting key genes.
scDR	[64]	scRNA-seq	Drug response scoring	Predicts drug response from scRNA-seq data and identifies resistance mechanisms.
scDrug	[90]	scRNA-seq	Bioinformatic workflow	Annotates tumor clusters, identifies DEGs, and predicts subpopulation drug response.
Single-cell analyses				
scBERT	[34]	scRNA-seq	Foundation model	Learns gene–gene interactions from large-scale scRNA-seq data for accurate cell type annotation and batch effect correction
scGen	[27]	scRNA-seq	VAEs	Combines VAEs and latent space vectors for high-dimensional single-cell gene expression data to accurately model perturbation.
scGPT	[35]	Single-cell multiomics	Foundation model	Offers flexible pipelines for multiomic integration, batch

Table 1 (*continued*)

Tool name	Ref.	Input data	Model/method	Description
scFoundation	[33]	scRNA-seq	Foundation model	correction, cell type annotation, perturbation prediction, and GRN inference. Provides several analyses, including gene expression enhancement, tissue and single-cell perturbation prediction, including drug response, cell type annotation and gene module inference.
scMAGeCK	[91]	scCRISPR screens	Robust rank aggregation	Identifies up- and downregulated genes upon gene knockout and constructs a genotype-phenotype network.
scDist	[17]	scRNA-seq	Mixed-effects model	Detects perturbed cell types while reducing false positives from individual and cohort variability.
Augur	[19]	scRNA-seq, STARmap, scATAC-seq	RF classifier	Prioritizes cell types based on their molecular response to a biological perturbation.
PerturbNet	[92]	scCRISPR screens	Deep generative model	Predicts the distribution of cell states induced by unseen chemical or genetic perturbations.
CPA	[52]	Single-cell and bulk RNA-Seq	AE	Prediction of combinatorial perturbation screens for drugs and genes
cycleCDR	[53]	Bulk RNA-seq, bulk proteomics, scRNA-seq	AE	Prediction of cellular responses to novel perturbations (e.g. unseen drugs)
CellOT	[54]	scRNA-seq	Optimal transport	Fully parameterized transport map to predict the cell state. Assumes that the perturbation induces subtle changes.
CellOracle	[22]	scRNA-seq	GRN modeling	Dissecting cell identity via network inference and in silico gene perturbation
MUSIC	[51]	CROP-seq, Perturb-Seq, CRISP-Seq	Topic modeling	Quantitative assessment of the perturbation impact on each gene
MIMOSCA	[3]	scRNA-seq	Perturbation analysis	Designs, performs, and analyzes perturbation scRNA-seq

Table 1 (*continued*)

Tool name	Ref.	Input data	Model/method	Description
				experiments to decode biological dynamics.
GEARS	[23]	scRNA-seq	Deep learning and knowledge graph	Predicts transcriptional responses to single and multigene perturbations using gene-gene relationships
trVAE	[29]	scRNA-seq	Transfer VAE	Predicts gene expression changes after perturbations through learning cell type-specific responses and improving out-of-distribution generalization
Spatial omics				
CELLama	[71]	scRNA-seq, ST	Language and the foundation model	Embeds multimodal data for applications such as cell typing and spatial context analysis.
Celcomen	[75]	scRNA-seq, ST	Generative graph neural network	Disentangles intra- and inter-cellular gene regulation to model perturbation effects in spatial and single-cell data
COMMOT	[69]	ST	Collective optimal transport	Infers cell-cell communication by considering ligand-receptor competition and spatial distances.
FlowSig	[64]	Single-cell transcriptomics or ST	Graphical causal modeling	Infers directed intercellular flows, from the input to the intracellular gene expression modules to the output of intercellular signals.
LIANA+	[68]	Single-cell and spatial omics	Multiview learning	Identifies inter- and intracellular signaling and CCC
MISTy	[62]	Multiplexed spatial data	Multiview learning	Extracts spatial omics relationships, such as GRNs and pathway activities.
SCING	[24]	Single-cell transcriptomics or ST	Gradient boosting and mutual information	Infers gene regulatory networks for physiology and disease insights.
SiGra	[70]	ST	Graph representation and the hybrid graph transformer	Enhances ST data to help uncover spatial architecture.
Experimental methods				
CROP-seq	[43]	Single-cell transcriptomics	Pooled CRISPR screens and droplet sequencing	Links gRNA expression to transcriptomic responses.
CRISP-seq	[44]	Single-cell transcriptomics	Pooled CRISPR screens	Combines CRISPR-pooled screens with scRNA-seq.

Table 1 (*continued*)

Tool name	Ref.	Input data	Model/method	Description
CUT&Tag	[40]	Chromatin profiling	Tagged transposase antibody	Examines histone modifications and transcription factors at single-cell resolution.
ECCITE-Seq	[46]	Single-cell transcriptomics and proteomics	Multimodal CRISPR screen	Captures multiomic data, integrating RNA and protein expression.
MIX-seq	[86]	scRNA-Seq	Multiplexed transcriptomics	Profiles perturbation responses across > 100 cell lines.
Perturb-CITE-Seq	[45]	Single-cell transcriptomics and proteomics	Pooled CRISPR screens	Integrates Perturb-Seq with proteomics using an antibody panel.
Perturb-Map	[63]	Multiplex imaging and ST	CRISPR with Pro-Codes	Analyzes gene knockout effects on tumor structure and immune infiltration.
Perturb-Seq	[3]	Single-cell transcriptomics	Pooled CRISPR screens	Assesses transcription factor effects using perturbation and scRNA-Seq.
sci-CAR	[41]	scRNA-seq and chromatin profiling	Single-cell combinatorial indexing with split-pool barcoding	Integrates RNA expression and chromatin accessibility in single cells.
scATAC-seq	[87]	Single-cell chromatin accessibility	Tn5 transposase tagging	Maps variations in regulatory element accessibility.
scBS-seq	[88]	Single-cell bisulfite sequencing	Bisulfite conversion	Measures DNA methylation in heterogeneous populations.
scTrio-seq	[38, 39]	Single-cell multiomics	RNA release and nuclei sequencing	Analyzes the transcriptome, methylome, and CNVs in the same cells.
TAP-seq	[89]	Single-cell transcriptomics	Genome-scale CRISPR screening	Maps enhancer-target gene interactions.

structures in scRNA-seq data [25]. Autoencoders (AEs) reduce dimensionality by encoding data into a lower dimensional space and reconstructing it, however, they often overfit, limiting generalizability [26, 27]. VAEs extend AE by introducing a distribution to the latent space [28]. This enables the model to learn representations of the input data. This generalizability allows VAEs to capture both condition- and cell type-specific information [27]. For instance, scGen applies a VAE framework to model single-cell perturbation responses across cell types, species, and experimental conditions [27]. This approach encodes unperturbed and perturbed single-cell gene expression into a latent space. A perturbation vector, representing the difference between the unperturbed and perturbed states, is calculated in this space. The decoder then reconstructs the corresponding gene expression profile, enabling scGen to predict perturbation responses for unseen conditions, such as dose-response effects and infection outcomes [27]. Meanwhile, transfer variational autoencoder extends the VAE framework to predict single-cell perturbation responses while addressing batch effects and domain adaptation challenges [29]. Similarly, the compositional perturbation autoencoder (CPA) builds on the VAE framework by modeling perturbation- and cell type-specific factors separately in the latent space [30]. CPA combines these factors to predict transcriptional

responses to unseen drug combinations, doses, or genetic interactions.

Foundation models based on transformer architectures have revolutionized scRNA-seq analysis through pretraining on diverse datasets to create scalable, generalized models [31,32]. These models capture complex transcriptional patterns, address batch effects, and integrate multiomics data [31,32]. By applying self-attention mechanisms and multihead attention layers, the transformer models capture dependencies across genes and cells. This ability to integrate information from thousands of genes makes these models exceptionally powerful tools for uncovering gene regulatory networks, cell-cell communication pathways, and cellular responses to perturbations (Fig. 2). Several foundation or large language models have recently been developed for single-cell analysis. Single-cell generative pretrained transformer (scGPT) addresses the challenges of pretraining on large-scale single-cell omics data using a transformer-based architecture trained on more than 33 million cells from 441 studies [33]. The framework has two stages: pretraining on large cell atlases to create general representations, then fine-tuning on smaller datasets for specific tasks [33]. scGPT uncovers cell type- and perturbation-specific biology while generalizing across studies and conditions. scFoundation employs an advanced encoder-decoder architecture with an asymmetric design, that was pretrained on more than 50 million cells spanning more than 100 tissues, cell types, and conditions. This model can be applied for half-maximal effective concentration (IC50) prediction, drug response modeling, and perturbation analysis [31]. Single-cell bidirectional encoder representations from transformers (scBERT) combines the BERT model architecture with gene embeddings to capture transcriptional patterns and gene-gene interactions [34]. Pretrained on large scRNA-seq datasets, scBERT specializes in cell type annotation, perturbation response prediction, and batch correction. Its self-attention mechanism enables the accurate identification of novel cell types and cross-cohort cell type assignments [34]. These foundation models provide scalable frameworks for advancing single-cell analysis, enabling accurate predictions of gene expression, perturbation responses, and cellular behavior across diverse conditions and datasets.

In summary, a range of analytical frameworks from classical models to advanced deep learning techniques have significantly improved our understanding of molecular responses to perturbations. Each approach has its strengths, and combining these methods provides a more comprehensive overview of cellular behavior.

3. Perturbation analysis of single-cell data

scRNA-seq has revolutionized biological research, offering an unprecedented level of detail in analyzing cellular responses to perturbations [35]. Despite the growing availability of single-cell omics datasets and resources such as the Human Cell Atlas [36], relatively few single-cell perturbational screens have been developed [15]. scRNA-seq enables researchers to understand the responses of cells to various stimuli with exceptional granularity. Its integration with other modalities, including single-cell proteomics and metabolomics, could expand its capacity to analyze complex cellular responses comprehensively. Combining scRNA-seq with genomic sequencing has provided a detailed understanding of the influence of copy number variations (CNVs) on gene expression at the single-cell level [37]. Furthermore, the integration of single-cell multiomics techniques has facilitated the exploration of epigenetic modifications with high resolution. Techniques such as scTrio-seq [38,39], single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq), single-cell bisulfite sequencing (scBS-seq), CUT&Tag [40], and sci-CAR [41] offer comprehensive insights into DNA methylation, histone modifications, and chromatin accessibility (Table 1). These advancements have provided a holistic view of the regulatory mechanisms driving cellular behavior, especially under perturbational conditions.

The most promising advancement in the field of genetic perturbation is single-cell CRISPR (scCRISPR) screening. This technique combines CRISPR-mediated gene editing with scRNA-seq to elucidate gene function, identify gene signatures, and evaluate cell states under diverse perturbational scenarios [42]. Perturb-Seq [3], CROP-seq [43] and CRISP-Seq [44] have enhanced scCRISPR screening (Fig. 1C). With guidance by CROP-seq in particular, RNA expression can be directly linked to transcriptome responses. Extensions such as Perturb-CITE-Seq [45] and ECCITE-Seq [46] incorporate proteomic data, providing additional layers of information about the impact of perturbations on both transcriptomic and proteomic levels by using selected antibodies to investigate proteins alongside gene expression data [47].

To support the analysis of scCRISPR screening data, advanced computational tools have been developed. For instance, the tool model-based understanding of single-cell CRISPR screening (MUSIC) [48] applies topic modeling to distinguish cellular perturbation states and identify critical regulatory changes. For each cell, the expression of each gene is analyzed and grouped into a number of topics required to distinguish the different perturbation states. Other computational models, such as CPA [30] and cycleCDR [49], have been designed to predict cellular responses to drug-induced or genetic perturbation. CPA leverages AE-based frameworks for this purpose, and they can handle experiments with multiple gene knockouts, thereby permitting the investigation of interactions between different genes and their combined effects on the cell. Meanwhile, cycleCDR applies two AEs to build a linear model for dynamic cellular behaviors. CellOT [50] uses the optimal transport theory to map subtle transitions between homeostatic and perturbed states, although its performance decreases with highly complex or strong perturbational effects. To improve the training of these models, several implementations have been developed to integrate perturbation data. For example, PerturBase [47] offers scRNA-seq and scATAC-seq data from 122 datasets (approximately 5 million cells) and provides tools to analyze genetic and chemical perturbations. PerturbDB [51] provides 66 Perturb-Seq datasets, featuring transcriptomes from 4.5 million cells and CRISPR-mediated knockdown for cancer research. scPerturb [52] harmonized 44 single-cell datasets for method benchmarking (Supplementary Table 1).

Despite the progress, significant challenges remain. Some computational tools might be less effective for highly complex perturbational responses, indicating a need for continued refinement. As the field advances, the combination of screening data with new multimodal models will improve our understanding of cellular mechanisms.

4. Dose-response analysis in single-cell and multiomics contexts

Single-cell methods have become powerful tools for high-throughput drug screening, enabling the prioritization of therapeutic options and the identification of potential drug synergies or resistance clusters within tumors. A key aspect of dose–response analysis is the estimation of the biological activity of a drug, often measured as the IC50, which represents the drug concentration required to reduce a given cellular response (e.g., growth, viability, or gene expression changes) by 50 % [53]. The approaches that model the effects of a drug at the single-cell level can capture these effective doses, key dose-dependent transcriptional responses and pathway perturbations [42,54].

By distinguishing sensitive and resistant cell populations, scRNA-seq enables the identification of the mechanisms underlying drug resistance and helps refine therapeutic strategies. This makes scRNA-seq ideal for identifying a drug's MoA, the causal biochemical and molecular pathways linking drug exposure to biological outcomes [42,55].

Based on their methodological approach, tools for dose-response analyses can be divided into bulk-to-single-cell transfer learning methods, which infer single-cell responses using bulk RNA-seq data, and direct single-cell prediction methods, which train models directly on single-cell data. Response modeling can be performed at the cluster or single-cell level. Bulk-to-single-cell transfer learning approaches use bulk RNA-seq drug-response datasets to train models that can then be applied to scRNA-seq data. This method relies on established bulk

databases, such as Genomics of Drug Sensitivity in Cancer (GDSC) [56] and Cancer Cell Line Encyclopedia (CCLE) [57], to train the models that require large amounts of data for the training process. Examples of tools employing this strategy include single-cell drug efficacy prediction with adversarial learning (scDEAL) [14], a deep transfer learning framework designed to predict drug responses at the single-cell level by integrating bulk and scRNA-seq data [14]. It employs two denoising AEs to extract low-dimensional gene features separately from bulk and scRNA-seq data [14]. Then, these representations are aligned through a domain-adaptive neural network. This approach ensures that drug response labels learned from bulk datasets can be effectively transferred to single-cell resolution. In addition, it applies integrated gradient interpretation to identify key gene signatures contributing to drug sensitivity or resistance [14]. Another approach to modeling drug responses is through perturbation-conditioned deep generative models, which learn transcriptional changes directly from perturbation data. PRnet is a deep generative model designed to predict gene expression responses to novel chemical perturbations at both the bulk and single-cell levels [13]. It integrates unperturbed transcriptional profiles and chemical perturbation data, and map them into a latent space that models how compounds influence gene activity [13]. The framework includes three core components: Perturb-adapter, which encodes chemical structures using SMILES representations, Perturb-encoder, which processes perturbation effects on baseline gene expression, and Perturb-decoder, which predicts transcriptional changes [13]. Another bioinformatics approach named drug response estimation from single-cell expression profiles (DREEP) is designed to predict drug sensitivity or resistance from scRNASeq data [58]. It integrates pharmacogenomic datasets, including GDSC, CTRP, and PRISM, to construct genomic profiles of drug sensitivity (GPDS)-ranked gene lists that capture associations between gene expression and drug potency [58]. Using gene set enrichment analysis (GSEA), DREEP compares single-cell expression profiles against GPDS signatures to the estimate drug response. A positive enrichment score indicates resistance-associated gene expression, whereas a negative score suggests sensitivity [58].

Unlike bulk-to-single-cell approaches, direct single-cell prediction methods operate exclusively on scRNA-seq data, capturing cell-specific responses and providing insights into cellular heterogeneity. For instance, scDR [59] predicts gene expression changes and drug-response scores to distinguish resistant and sensitive cell types across cancer types. Additional tools such as scDRUG [90] and Beyondcell [83] aggregate data from cell populations to reduce noise and improve computational efficiency. The most comprehensive repositories for drug responses are PRISM, CCLE, and GDSC [56,57,60], which provide extensive information on drug sensitivity and gene expression in cell lines mostly related to cancer. Despite newer developments such as PerturBase [47] and CeDR Atlas [61], which compile single-cell drug response data, no truly comprehensive resource that captures all necessary aspects for in-depth modeling exists. A unified, extensive database that covers a wide range of cell types, perturbation conditions, and dose-response relationships is missing. Supplementary Table 1 summarizes some of the largest currently available single-cell drug response studies, but more effort is needed to develop an all-encompassing resource that can support more reliable predictions and cross-study analyses.

Overall, single-cell methods enable high-throughput drug screening and detailed exploration of drug MoA. The continued development of these tools and resources will advance our understanding of therapeutic heterogeneity, identify resistance mechanisms, and uncover novel therapeutic targets in oncology research.

5. Spatial approaches for perturbation modeling

scRNA-seq data has several advantages in perturbation analyses, including the possibility of deciphering different cell types and stages. However, spatial contexts, which are critical for understanding tissue architecture and cell-cell interactions, are lost [62]. Spatial transcriptomics (ST) addresses this limitation by preserving spatial information while profiling gene expression. Although ST has primarily been used to analyze tissue organization, it is increasingly being adapted to investigate cellular responses to perturbations. One example is Perturb-Map, which employs protein barcoding (Pro-Codes) to mark cells expressing CRISPR gRNAs. These Pro-Codes can be used to assess the effects of gene knockout on tumor architecture and immune cell infiltration through multiplex imaging and ST [63]. However, the application of ST to perturbational studies remains limited, partly because of the lack of paired ST data before and after perturbation [64]. Consequently, tools that are not specifically designed for perturbation screens but can be adapted to compare control and perturbed states in different scenarios have been applied (Table 1). One of the advantages of ST in perturbation biology is hypothesis generation. The large datasets produced by ST permit the identification of novel ligand-pathway interactions, as demonstrated by tools such as multiview intercellular spatial modeling (MISTy). MISTy uses explainable ML on multiplexed spatial technologies, such as ST and imaging mass cytometry, to understand marker interactions by profiling intra- and intercellular relationships. It constructs models to describe different spatial contexts, including the intrinsic, local, and tissue views [62]. Other tools such as PROGENy [65] and OmniPath [66] can be included to add information about signaling pathway activity and ligand gene expression, respectively [62]. By integrating such tools, researchers can better understand the influence of perturbations on signaling pathways and cellular communication. Furthermore, perturbation can alter cell-cell communication (CCC), which can be analyzed using transcriptomic data by extracting information from receptor-ligand interaction databases such as CellPhoneDB [67]. LIANA+ extends MISTy's multiview approach, enabling the joint modeling of combinations of complex tissue structures and functions. It can infer receptor-ligand interactions from ST data in an unsupervised manner or from scRNA-seq data across multiple conditions [68]. Methods such as communication analysis by optimal transport (COMMOT) extend this approach by modeling the directionality of CCC in ST data, which is essential for understanding the dynamic signaling response to perturbations. This is made possible by using collective optimal transport to infer CCC in space by introducing the distributions of ligands and receptors and enforcing spatial constraints [69]. However, a limitation of transcriptomics-based CCC inference is the lack of direct protein-level data, leading to potential false positives [69]. In addition, FlowSig allows the inference of intercellular flows using ST or scRNA-seq data. It employs causal modeling to construct directed graphs and identify signal flow and is sufficient to map inflowing intercellular signals (inputs) to gene expression modules that mediate intracellular responses and generate outflowing intercellular signals (outputs) at each spatial location [64]. These graph-based approaches are intuitive for ST, as the spatial relationships between cells can be naturally represented [70]. They can also help to improve the analysis of ST data, as demonstrated by single-cell spatial elucidation through image-augmented graph transformer (SiGra), because they usually have a low total number of transcripts per cell, noisy data and significant zeros. SiGra enhances gene expression data by combining single-cell and spatial multimodal data, including imaging and transcriptomics, into a graphical model with three graph transformer-based encoder-decoders. This enhanced gene expression facilitates insights into CCC and other biological discoveries [55]. Both spatial and single-cell information can be captured through transcriptomics, as well as genomics, epigenomics, proteomics, and metabolomics, providing a more comprehensive picture of the effects of perturbations on cellular systems. However, these different types of data must be integrated to obtain information about perturbational changes. Emerging ML frameworks, such as foundation models, can integrate different types of data modalities. For example, cell embedding leverage language model abilities (CELLama) embeds multimodal data into a shared space, enabling the integration of scRNA-seq, ST data, and metadata for

enhanced predictions [71]. SCING can also integrate multiomics data, including ST data and protein expression data from immunohistochemistry. Using scRNA-seq or ST, SCING infers GRNs, which are crucial for perturbation analysis, as they reveal changes in gene-gene interactions [24]. Despite these advances, ST and scRNA-seq face barriers

to clinical translation, including high costs and technical complexity. To address this, tools have emerged to predict gene expression from the frequently used imaging technique of hematoxylin and eosin (H&E) staining (e.g., HE2RNA, Hist2ST [72,73]). Furthermore, tools such as tumor edge structure and lymphocyte multi-level annotation (TESLA)
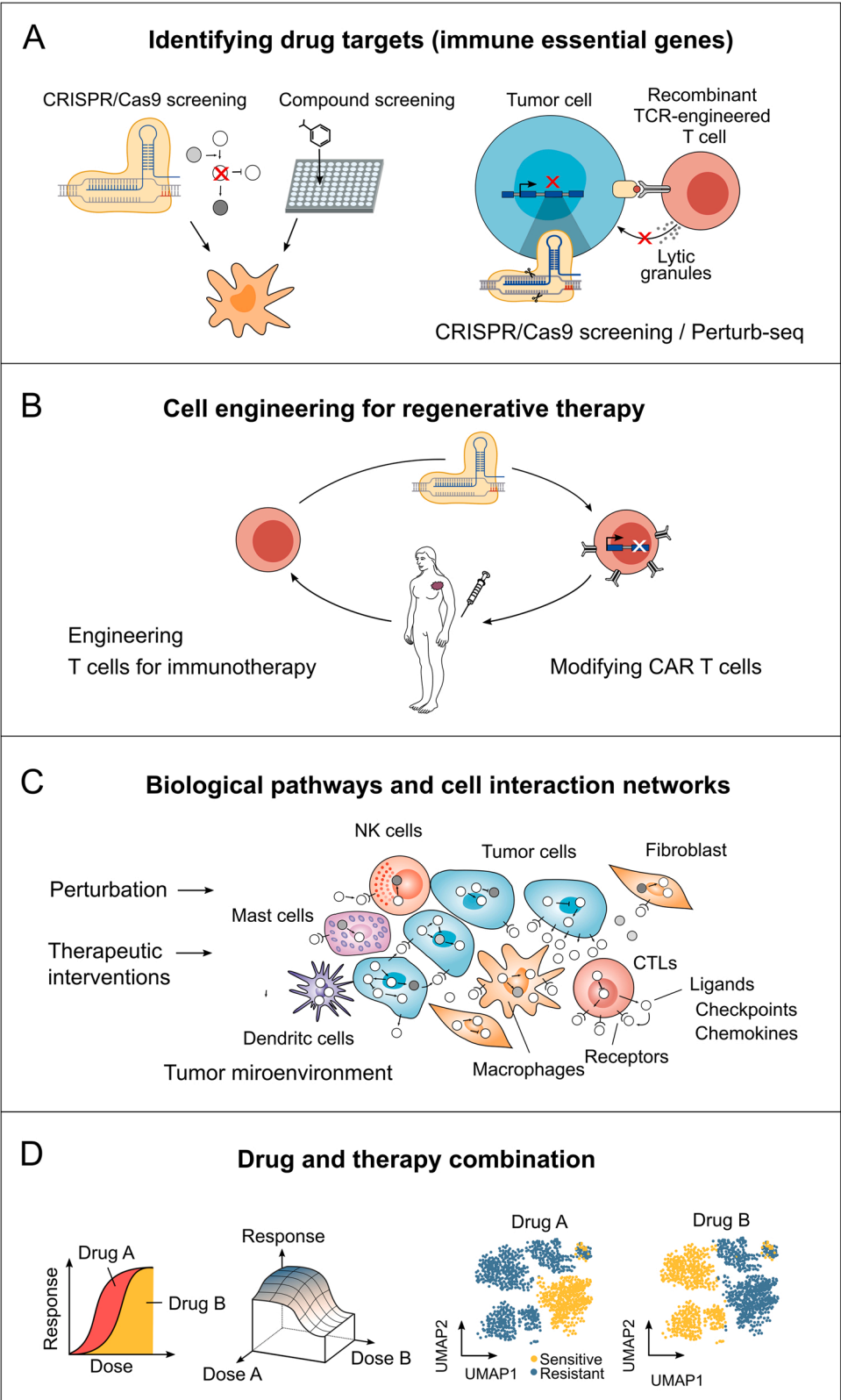

Fig. 3. Applications for perturbation modeling and screening.

can use H&E stainings to enhance the spatial resolution of ST for the tumor region annotation [74]. In the future, such techniques could be used in clinical contexts to infer perturbation effects at the tissue level.

In summary, although currently available tools for spatial analyses are not explicitly designed for perturbation studies, they can be adapted to address perturbation-related questions, including hypothesis generation or CCC inference. One recent approach named Celcomen [75] focused on spatial causal disentanglement for single-cell and tissue perturbation modeling through a generative graphical neural network. It is anticipated that novel computational tools will be developed to specifically address the integration challenges associated with perturbation biology.

6. Applications

Technologies such as CRISPR/Cas9 and RNAi can be used to introduce various genetic manipulations that lead to changes in the function or expression of certain genes. These methods are crucial in screening for genetic modifications that result in specific phenotypic or functional changes in cells, potentially facilitating the identification of drug targets. For instance, screening of cancer cell lines can facilitate the development of targeted therapies. Moreover, these techniques are highly valuable for immunotherapy. By using a two-cell co-culture model that systematically modifies genes in tumor cells and reads out the cytotoxic function of T cells, researchers can identify immune-essential genes (Fig. 3A). This could even be improved when combined with scRNA-seq techniques, such as Perturb-Seq or CROP-seq to enable large-scale perturbation studies with transcriptional readouts.

Another promising application relevant to immunotherapy and regenerative cell therapy is to identify the optimal manipulation for engineered T cells (or CAR T cells, Fig. 3B). In general, pooled genetic perturbation screens are extremely powerful, and they can be applied in many applications, allowing the comprehensive characterization of cell models such as organoids. Another emerging technology is optical pooled screens, which use imaging or in situ sequencing or hybridization as readouts [1]. In addition, spatially resolved transcriptomics provides a detailed view of gene expression in tissue contexts. Although challenges such as sensitivity limitations and labor-intensive protocols remain [76], with recent developments involving the use of RNA imaging-based techniques (MERFISH) or the padlock-based Xenium in situ platform (10X), a readout at the single cell resolution of thousands of genes has become possible. This expands the scope of our analyses making it possible to investigate the interaction between the same cell types, but also between different cell types, and how they can be affected by certain perturbations or treatments. The tumor microenvironment in solid tumors represents an ideal application, in which the communication and co-localization between tumor cells and a number of other cell types such as cytotoxic T lymphocytes and changes in their patterns by certain interventions can be investigated (Fig. 3C). New computational tools, such as the spatial foundation model CELLama [71], are being developed to integrate spatial data with single-cell information aiming to enhance spatial context analysis and perturbation prediction. Interestingly, a spatial CRISPR screen (PerturbMap) show its ability to identify tumor specific modulation of macrophages in the tumor microenvironment in ovarian cancer [77]

From cancer treatment studies, an increasing number of scRNA-seq data, as outlined in Supplementary Table 1, are available. If systematically acquired, analyzed, and integrated into perturbation atlases, these datasets could contribute to our understanding of resistance mechanisms. The availability of other data modalities is limited; therefore, it is difficult to address whether the integration of other readouts can improve characterization of drug responses [15]. Perturbation modeling could also be used to investigate drug and therapy combinations and their responses. One possibility to expand genetic screens for target identification is analyzing synthetic lethal interactions using simultaneous perturbations of two genes resulting in cell death or combining

gene perturbation with small molecules or other compounds that can be used to elucidate the MoAs of drugs. In any case, the response to drug combinations can also be investigated directly by interaction modeling. For multicellular systems additional readouts from scRNA-seq are extremely valuable for identifying mutually resistant cell populations and prioritizing drug combinations (Fig. 3D). Single-cell analyses would also be relevant in investigating drug repurposing and dose-response analyses to determine effective concentrations and toxicologic thresholds. However, few single-cell datasets are available considering the response at different concentrations.

7. Conclusions and outlook

The promise of foundation models including scFoundation [31] lies in their ability to enhance our understanding of cellular and tissue-level biology. However, their progress requires continuous innovation in the computational approaches and even more importantly huge amounts of training data. The development of comprehensive resources such as single cell atlases including tumor tissues [78,79] has also sparked the generation of perturbation cell and tissue atlases [1]. PerturbAtlas and PerturbDB represent such platforms, which systematically map gene functions and the effects of genetic perturbations across various cell types, accelerating discoveries, and providing benchmarks for model evaluation. Specifically for tissue-level studies, spatial approaches such as PerturbMap are promising, but there is a lack of paired spatial transcriptomics data before and after perturbation, so the number of approaches for perturbation analysis in spatial settings remains limited. With increasing availability of ST data, further computational approaches will be adapted or new ones developed to reveal changes in the spatial cell distribution, which will ultimately improve the situation for perturbation modeling in this context.

Benchmarking efforts such as those outlined in Therapeutics Data Common, which provides ML datasets for drug discovery and development [80] should prompts efforts for standardized and accessible data or tasks. We anticipate a growing number of ML models for perturbation analyses with similar impact as we observed for protein structure predictions [81].In conclusion, recent developments in AI/ML as well as emerging single cell technologies hold great promise for perturbation modeling to improve drug development and personalized medicine. However, continued collaboration between experimental and computational biologists will be critical to overcome current limitations and develop tools that are both predictive and interpretable. The combination of foundation models and perturbation atlases offers immense potential for advancing our understanding of cellular behavior and disease mechanisms.

CRediT authorship contribution statement

Monfort-Lanzas Pablo: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Hackl Hubert:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Madersbacher Leonie:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation. **Rungger Katja:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL and chatGPT-4.0 in order to improve readability and language. After using this tool/services, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Declaration of Competing Interest

Acknowledgements

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.02.028.

References

[1] Rood JE, Hupalowska A, Regev A. Toward a foundation model of causal cell and tissue biology with a perturbation cell and tissue atlas. Cell 2024;187:4520–45. https://doi.org/10.1016/j.cell.2024.07.035.

[2] Gavriilidis GI, Vasileiou V, Orfanou A, Ishaque N, Psomopoulos F. A mini-review on perturbation modelling across single-cell omic modalities. Comput Struct Biotechnol J 2024;23:1886–96. https://doi.org/10.1016/j.csbj.2024.04.058.

[3] Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell 2016;167:1853–66. https://doi.org/10.1016/j.cell.2016.11.038 [.e17].

[4] Sarmah D, Smith GR, Bouhaddou M, Stern AD, Erskine J, Birtwistle MR. Network inference from perturbation time course data. Npj Syst Biol Appl 2022;8:42. https://doi.org/10.1038/s41540-022-00253-6.

[5] Noh H, Shoemaker JE, Gunawan R. Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection. Nucleic Acids Res 2018;46. https://doi.org/10.1093/nar/gkx1314 [e34–e34].

[6] Molinelli EJ, Korkut A, Wang W, Miller ML, Gauthier NP, Jing X, et al. Perturbation biology: inferring signaling networks in cellular systems. PLoS Comput Biol 2013;9: e1003290. https://doi.org/10.1371/journal.pcbi.1003290.

[7] Eduati F, Doldàn-Martelli V, Klinger B, Cokelaer T, Sieber A, Kogera F, et al. Drug resistance mechanisms in colorectal cancer dissected with cell type–specific dynamic logic models. Cancer Res 2017;77:3364–75. https://doi.org/10.1158/0008-5472.CAN-17-0078.

[8] Gjerga E, Trairatphisan P, Gabor A, Koch H, Chevalier C, Ceccarelli F, et al. Converting networks to predictive logic models from perturbation signalling data with CellNOpt. Bioinformatics 2020;36:4523–4. https://doi.org/10.1093/bioinformatics/btaa561.

[9] Abou-Jaoudé W, Traynard P, Monteiro PT, Saez-Rodriguez J, Helikar T, Thieffry D, et al. Logical modeling and dynamical analysis of cellular networks. Front Genet 2016:7. https://doi.org/10.3389/fgene.2016.00094.

[10] Niktab E, Atkinson PH, Walterfang M, Winship I, Sturley SL, Munkacsi AB. Ordinary differential equations to construct invertible generative models of cell type and tissue-specific regulatory networks; 2023. ⟨https://doi.org/10.1101/2023.05.18.540731⟩.

[11] Qian L, Sun R, Aebersold R, Bühlmann P, Sander C, Guo T. AI-empowered perturbation proteomics for complex biological systems. Cell Genom 2024;4: 100691. https://doi.org/10.1016/j.xgen.2024.100691.

[12] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 2017;171:1437–52. https://doi.org/10.1016/j.cell.2017.10.049 [.e17].

[13] Qi X, Zhao L, Tian C, Li Y, Chen Z-L, Huo P, et al. Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. Nat Commun 2024;15:9256. https://doi.org/10.1038/s41467-024-53457-1.

[14] Chen J, Wang X, Ma A, Wang Q-E, Liu B, Li L, et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. Nat Commun 2022;13:6494. https://doi.org/10.1038/s41467-022-34277-7.

[15] Ji Y, Lotfollahi M, Wolf FA, Theis FJ. Machine learning for perturbational single-cell omics. Cell Syst 2021;12:522–37. https://doi.org/10.1016/j.cels.2021.05.016.

[16] Li L, You Y, Liao W, Fan X, Lu S, Cao Y, et al. A systematic comparison of single-cell perturbation response prediction models; 2024. ⟨https://doi.org/10.1101/2024.12.23.630036⟩.

[17] Nicol PB, Paulson D, Qian G, Liu XS, Irizarry R, Sahu AD. Robust identification of perturbed cell types in single-cell RNA-seq data. Nat Commun 2024;15:7610. https://doi.org/10.1038/s41467-024-51649-3.

[18] Skinnider MA, Squair JW, Kathe C, Anderson MA, Gautier M, Matson KJE, et al. Cell type prioritization in single-cell data. Nat Biotechnol 2021;39:30–4. https://doi.org/10.1038/s41587-020-0605-1.

[19] Kana O, Nault R, Filipovic D, Marri D, Zacharewski T, Bhattacharya S. Generative modeling of single-cell gene expression for dose-dependent chemical perturbations. Patterns 2023;4:100817. https://doi.org/10.1016/j.patter.2023.100817.

[20] Momeni K, Ghorbian S, Ahmadpour E, Sharifi R. Unraveling the complexity: understanding the deconvolutions of RNA-seq data. Transl Med Commun 2023;8: 21. https://doi.org/10.1186/s41231-023-00154-8.

[21] Badia-i-Mompel P, Wessels L, Müller-Dott S, Trimbour R, Ramirez Flores RO, Argelaguet R, et al. Gene regulatory network inference in the era of single-cell multi-omics. Nat Rev Genet 2023;24:739–54. https://doi.org/10.1038/s41576-023-00618-5.

[22] Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature 2023;614:742–51. https://doi.org/10.1038/s41586-022-05688-9.

[23] Roohani Y, Huang K, Leskovec J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. Nat Biotechnol 2024;42:927–35. https://doi.org/10.1038/s41587-023-01905-6.

[24] Littman R, Cheng M, Wang N, Peng C, Yang X. SCING: inference of robust, interpretable gene regulatory networks from single cell and spatial transcriptomics. iScience 2023;26:107124. https://doi.org/10.1016/j.isci.2023.107124.

[25] Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun 2018;9:2002. https://doi.org/10.1038/s41467-018-04368-5.

[26] Monfort-Lanzas P, Gronauer R, Balaz M, Floriani G, Hackl H. Application of artificial intelligence in immuno-oncology. Ref. module biomed. sci. Elsevier; 2024. https://doi.org/10.1016/B978-0-443-14064-8.00017-5.

[27] Lotfollahi M, Wolf FA, Theis FJ. scGen predicts single-cell perturbation responses. Nat Methods 2019;16:715–21. https://doi.org/10.1038/s41592-019-0494-8.

[28] Kingma DP, Welling M. Auto-encoding variational bayes; 2013. ⟨https://doi.org/10.48550/ARXIV.1312.6114⟩.

[29] Lotfollahi M, Naghipourfar M, Theis FJ, Wolf FA. Conditional out-of-distribution generation for unpaired data using transfer VAE. Bioinformatics 2020;36:i610–7. https://doi.org/10.1093/bioinformatics/btaa800.

[30] Lotfollahi M, Klimovskaia Susmelj A, De Donno C, Hetzel L, Ji Y, Ibarra IL, et al. Predicting cellular responses to complex perturbations in high-throughput screens. Mol Syst Biol 2023;19:e11517. https://doi.org/10.15252/msb.202211517.

[31] Hao M, Gong J, Zeng X, Liu C, Guo Y, Cheng X, et al. Large-scale foundation model on single-cell transcriptomics. Nat Methods 2024;21:1481–91. https://doi.org/10.1038/s41592-024-02305-7.

[32] Boiarsky R, Singh N, Buendia A, Getz G, Sontag D. A deep dive into single-cell RNA sequencing foundation models; 2023. ⟨https://doi.org/10.1101/2023.10.19.563100⟩.

[33] Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods 2024. https://doi.org/10.1038/s41592-024-02201-0.

[34] Yang F, Wang W, Wang F, Fang Y, Tang D, Huang J, et al. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nat Mach Intell 2022;4:852–66. https://doi.org/10.1038/s42256-022-00534-z.

[35] Svensson V, Da Veiga Beltrame E, Pachter L. A curated database reveals trends in single-cell transcriptomics. Database 2020;2020:baaa073. https://doi.org/10.1093/database/baaa073.

[36] Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas. eLife 2017;6:e27041. https://doi.org/10.7554/eLife.27041.

[37] Dey SS, Kester L, Spanjaard B, Bienko M, Van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol 2015;33:285–9. https://doi.org/10.1038/nbt.3129.

[38] Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res 2016;26:304–19. https://doi.org/10.1038/cr.2016.23.

[39] Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol 2023;24:695–713. https://doi.org/10.1038/s41580-023-00615-w.

[40] Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 2019;10:1930. https://doi.org/10.1038/s41467-019-09982-5.

[41] Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 2018;361:1380–5. https://doi.org/10.1126/science.aau0730.

[42] 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;22:496–520. https://doi.org/10.1038/s41573-023-00688-4.

[43] 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;14:297–301. https://doi.org/10.1038/nmeth.4177.

[44] Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. Cell 2016;167:1883–96. https://doi.org/10.1016/j.cell.2016.11.039 [.e15].

[45] Frangieh CJ, Melms JC, Thakore PI, Geiger-Schuller KR, Ho P, Luoma AM, et al. Multimodal pooled perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. Nat Genet 2021;53:332–41. https://doi.org/10.1038/s41588-021-00779-1.

[46] Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods 2019;16:409–12. https://doi.org/10.1038/s41592-019-0392-0.

[47] Wei Z, Si D, Duan B, Gao Y, Yu Q, Zhang Z, et al. PerturBase: a comprehensive database for single-cell perturbation data analysis and visualization. Nucleic Acids Res 2024:gkae858. https://doi.org/10.1093/nar/gkae858.

[48] Duan B, Zhou C, Zhu C, Yu Y, Li G, Zhang S, et al. Model-based understanding of single-cell CRISPR screening. Nat Commun 2019;10:2233. https://doi.org/10.1038/s41467-019-10216-x.

[49] Huang W, Liu H. Predicting single-cell cellular responses to perturbations using cycle consistency learning. Bioinformatics 2024;40:i462–70. https://doi.org/10.1093/bioinformatics/btae248.

[50] Bunne C, Stark SG, Gut G, Del Castillo JS, Levesque M, Lehmann K-V, et al. Learning single-cell perturbation responses using neural optimal transport. Nat Methods 2023;20:1759–68. https://doi.org/10.1038/s41592-023-01969-x.

[51] Yang B, Zhang M, Shi Y, Zheng B-Q, Shi C, Lu D, et al. PerturbDB for unraveling gene functions and regulatory networks. Nucleic Acids Res 2025;53:D1120–31. https://doi.org/10.1093/nar/gkae777.

[52] Peidli S, Green TD, Shen C, Gross T, Min J, Garda S, et al. scPerturb: harmonized single-cell perturbation data. Nat Methods 2024;21:531–40. https://doi.org/10.1038/s41592-023-02144-y.

[53] Ewald JD, Basu N, Crump D, Boulanger E, Head J. Characterizing variability and uncertainty associated with transcriptomic dose–response modeling. Environ Sci Technol 2022;56:15960–8. https://doi.org/10.1021/acs.est.2c04665.

[54] Gezelius H, Enblad AP, Lundmark A, Åberg M, Blom K, Rudfeldt J, et al. Comparison of high-throughput single-cell RNA-seq methods for ex vivo drug screening. NAR Genom Bioinform 2024;6:lqae001. https://doi.org/10.1093/nargab/lqae001.

[55] Kinaret PAS, Serra A, Federico A, Kohonen P, Nymark P, Liampa I, et al. Transcriptomics in toxicogenomics, Part I: experimental design, technologies, publicly available data, and regulatory aspects. Nanomaterials 2020;10:750. https://doi.org/10.3390/nano10040750.

[56] Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 2012;41:D955–61. https://doi.org/10.1093/nar/gks1111.

[57] Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the cancer cell line encyclopedia. Nature 2019;569:503–8. https://doi.org/10.1038/s41586-019-1186-3.

[58] Pellecchia S, Viscido G, Franchini M, Gambardella G. Predicting drug response from single-cell expression profiles of tumours. BMC Med 2023;21:476. https://doi.org/10.1186/s12916-023-03182-1.

[59] Lei W, Yuan M, Long M, Zhang T, Huang Y, Liu H, et al. scDR: Predicting drug response at single-cell resolution. Genes 2023;14:268. https://doi.org/10.3390/genes14020268.

[60] Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer 2020;1:235–48. https://doi.org/10.1038/s43018-019-0018-6.

[61] Wang Y-Y, Kang H, Xu T, Hao L, Bao Y, Jia P. CeDR atlas: a knowledgebase of cellular drug response. Nucleic Acids Res 2022;50:D1164–71. https://doi.org/10.1093/nar/gkab897.

[62] Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. Genome Biol 2022;23:97. https://doi.org/10.1186/s13059-022-02663-5.

[63] Dhainaut M, Rose SA, Akturk G, Wroblewska A, Nielsen SR, Park ES, et al. Spatial CRISPR genomics identifies regulators of the tumor microenvironment. Cell 2022;185:1223–39. https://doi.org/10.1016/j.cell.2022.02.015 [.e20].

[64] Almet AA, Tsai Y-C, Watanabe M, Nie Q. Inferring pattern-driving intercellular flows from single-cell and spatial transcriptomics. Nat Methods 2024;21:1806–17. https://doi.org/10.1038/s41592-024-02380-w.

[65] Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun 2018;9:20. https://doi.org/10.1038/s41467-017-02391-6.

[66] Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods 2016;13:966–7. https://doi.org/10.1038/nmeth.4077.

[67] Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc 2020;15:1484–506. https://doi.org/10.1038/s41596-020-0292-x.

[68] Dimitrov D, Schäfer PSL, Farr E, Rodriguez-Mier P, Lobentanzer S, Badia-i-Mompel P, et al. LIANA+ provides an all-in-one framework for cell–cell communication inference. Nat Cell Biol 2024;26:1613–22. https://doi.org/10.1038/s41556-024-01469-w.

[69] Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. Nat Methods 2023;20:218–28. https://doi.org/10.1038/s41592-022-01728-4.

[70] Tang Z, Li Z, Hou T, Zhang T, Yang B, Su J, et al. SiGra: single-cell spatial elucidation through an image-augmented graph transformer. Nat Commun 2023;14:5618. https://doi.org/10.1038/s41467-023-41437-w.

[71] Choi H, Park J, Kim S, Kim J, Lee D, Bae S, et al. CELLama: foundation model for single cell and spatial transcriptomics by cell embedding leveraging language model abilities; 2024. ⟨https://doi.org/10.1101/2024.05.08.593094⟩.

[72] Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun 2020;11:3877. https://doi.org/10.1038/s41467-020-17678-4.

[73] Zeng Y, Wei Z, Yu W, Yin R, Yuan Y, Li B, et al. Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. Brief Bioinform 2022;23:bbac297. https://doi.org/10.1093/bib/bbac297.

[74] Hu J, Coleman K, Zhang D, Lee EB, Kadara H, Wang L, et al. Deciphering tumor ecosystems at super resolution from spatial transcriptomics with TESLA. Cell Syst 2023;14:404–17. https://doi.org/10.1016/j.cels.2023.03.008 [.e4].

[75] Megas S, Chen DG, Polanski K, Eliasof M, Schonlieb C-B, Teichmann SA. Celcomen: spatial causal disentanglement for single-cell and tissue perturbation modeling; 2024. ⟨https://doi.org/10.48550/ARXIV.2409.05804⟩.

[76] Asp M, Bergenstråhle J, Lundeberg J. Spatially resolved transcriptomes—next generation tools for tissue exploration. BioEssays 2020;42:1900221. https://doi.org/10.1002/bies.201900221.

[77] Mollaoglu G, Tepper A, Falcomatà C, Potak HT, Pia L, Amabile A, et al. Ovarian cancer-derived IL-4 promotes immunotherapy resistance. Cell 2024;187:7492–510. https://doi.org/10.1016/j.cell.2024.10.006 [.e22].

[78] Salcher S, Sturm G, Horvath L, Untergasser G, Kuempers C, Fotakis G, et al. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. Cancer Cell 2022;40:1503–20. https://doi.org/10.1016/j.ccell.2022.10.008 [.e8].

[79] Marteau V, Nemati N, Handler K, Raju D, Soto EK, Fotakis G, et al. Single-cell integration and multi-modal profiling reveals phenotypes and spatial organization of neutrophils in colorectal cancer; 2024. ⟨https://doi.org/10.1101/2024.08.26.609563⟩.

[80] Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, et al. Therapeutics data commons: machine learning datasets and tasks for drug discovery and development; 2021. ⟨https://doi.org/10.48550/ARXIV.2102.09548⟩.

[81] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024;630:493–500. https://doi.org/10.1038/s41586-024-07487-w.

[82] He B, Xiao Y, Liang H, Huang Q, Du Y, Li Y, et al. ASGARD is A single-cell guided pipeline to aid repurposing of drugs. Nat Commun 2023;14:993. https://doi.org/10.1038/s41467-023-36637-3.

[83] 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;13:187. https://doi.org/10.1186/s13073-021-01001-x.

[84] Suphavilai C, Chia S, Sharma A, Tu L, Da Silva RP, Mongia A, et al. Predicting heterogeneity in clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. Genome Med 2021;13:189. https://doi.org/10.1186/s13073-021-01000-y.

[85] Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. Cancer Cell 2020;38:672–84. https://doi.org/10.1016/j.ccell.2020.09.014 [.e6].

[86] McFarland JM, Paolella BR, Warren A, Geiger-Schuller K, Shibue T, Rothberg M, et al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. Nat Commun 2020;11:4296. https://doi.org/10.1038/s41467-020-17440-w.

[87] Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486–90. https://doi.org/10.1038/nature14590.

[88] Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 2014;11:817–20. https://doi.org/10.1038/nmeth.3035.

[89] Schraivogel D, Gschwind AR, Milbank JH, Leonce DR, Jakob P, Mathur L, et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. Nat Methods 2020;17:629–35. https://doi.org/10.1038/s41592-020-0837-5.

[90] Hsieh C-Y, Wen J-H, Lin S-M, Tseng T-Y, Huang J-H, Huang H-C, et al. scDrug: From single-cell RNA-seq to drug response prediction. Comput Struct Biotechnol J 2023;21:150–7. https://doi.org/10.1016/j.csbj.2022.11.055.

[91] Yang L, Zhu Y, Yu H, Cheng X, Chen S, Chu Y, et al. scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biol 2020;21:19. https://doi.org/10.1186/s13059-020-1928-4.

[92] H Yu JD Welch PerturbNet predicts single-cell responses to unseen chemical and genetic perturbations 2022.⟨https://doi.org/10.1101/2022.07.20.500854⟩.