Computational and Structural Biotechnology Journal 23 (2024) 1886–1896

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Mini-review

A mini-review on perturbation modelling across single-cell omic modalities

George I. Gavriilidis [a,*], Vasileios Vasileiou [a,b], Aspasia Orfanou [a], Naveed Ishaque [c], Fotis Psomopoulos [a,*]

[a] Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece
[b] Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece
[c] Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Center of Digital Health, Berlin, Germany

ARTICLE INFO

Keywords:
Perturbation
Single-cell RNA sequencing
ScRNAseq
Machine learning
Deep learning

ABSTRACT

Recent advances in single-cell omics technology have transformed the landscape of cellular and molecular research, enriching the scope and intricacy of cellular characterisation. Perturbation modelling seeks to comprehensively grasp the effects of external influences like disease onset or molecular knock-outs or external stimulants on cellular physiology, specifically on transcription factors, signal transducers, biological pathways, and dynamic cell states. Machine and deep learning tools transform complex perturbational phenomena in algorithmically tractable tasks to formulate predictions based on various types of single-cell datasets. However, the recent surge in tools and datasets makes it challenging for experimental biologists and computational scientists to keep track of the recent advances in this rapidly expanding filed of single-cell modelling. Here, we recapitulate the main objectives of perturbation modelling and summarise novel single-cell perturbation technologies based on genetic manipulation like CRISPR or compounds, spanning across omic modalities. We then concisely review a burgeoning group of computational methods extending from classical statistical inference methodologies to various machine and deep learning architectures like shallow models or autoencoders, to biologically informed approaches based on gene regulatory networks, and to combinatorial efforts reminiscent of ensemble learning. We also discuss the rising trend of large foundational models in single-cell perturbation modelling inspired by large language models. Lastly, we critically assess the challenges that underline single-cell perturbation modelling while pointing towards relevant future perspectives like perturbation atlases, multi-omics and spatial datasets, causal machine learning for interpretability, multi-task learning for performance and explainability as well as prospects for solving interoperability and benchmarking pitfalls.

1. Introduction

Emerging paradigms in single-cell and spatial multi-modal omics have revolutionised the resolution at which cellular and molecular research is conducted. While these technologies increase the phenotypic depth and breadth of cellular characterisation, there is also a need to model dynamic behaviour in these systems, such as measuring and predicting the effect of perturbations caused by external influences (e.g., molecular knockouts, over expressions, application of drugs, or temperature). A previous review in the field aptly coined the term **"perturbation modelling"** to refer to these innovative single-cell studies with significant pharmacological and translation potential [1].

The overarching purpose of perturbation modelling is to provide a holistic view of how external biological or chemical interventions can perturb aspects of cell physiology, such as transcription factors, signal transducers, biological pathways, and dynamic cell states. These perturbations can be measured and predicted across time, micro-anatomical regions, tissues, or even organs. Modelling perturbation response is crucial to designing novel therapeutic interventions and studying disease progression, making it ideal for patient-tailored approaches. Perturbation modelling is underpinned by systematic databases of biological response to treatment (e.g. Connectivity Map [2], L1000, DRUG-seq [3]) and leverages the significant potential of machine learning (ML) to make predictive models, e.g., molecular targets or IC50 values. Furthermore, deep learning (DL) has become a focal point of perturbation modelling since it is more suitable for unravelling complex, non-linear dynamics across multiple omic modalities.

Over the last couple of years, the field has seen a meteoric rise in the number of perturbation modelling tools for various single-cell technologies. Projections show that this type of analysis will become part of the

* Corresponding authors.
E-mail addresses: ggeorav@certh.gr (G.I. Gavriilidis), fpsom@certh.gr (F. Psomopoulos).

https://doi.org/10.1016/j.csbj.2024.04.058
Received 1 February 2024; Received in revised form 23 April 2024; Accepted 23 April 2024
Available online 25 April 2024

typical and standard single-cell pipelines in the coming years. Even so, their complex nature, which is a cross-fertilisation of pharmacology, biology, mathematics, and computer science, along with a lack of standards, metrics, and extensive benchmarks, hinders the broader adoption of these tools by the scientific community.

This mini-review aims to provide a concise overview of recent perturbation modelling techniques. It focuses on the basic principles and strategies and provides examples of tools and software that use different approaches. We also briefly discuss challenges and promising future directions for method development and application.

2. Basic principles of perturbation modelling

Based on the seminal paper by Ji et al. [4], four cardinal objectives (O1-O4) can be outlined for perturbation modelling that traverses single-cell datasets and methodologies, addressed below.

A primary objective of single-cell perturbation modelling is to "*extrapolate and elucidate*" perturbations **(O1)**. More concretely, this objective partly entails models that *extrapolate* to unseen omic and phenotypic changes a cell can experience due to a perturbation, often referred to as *out-of-distribution* (OOD) detection. Some of these models try to predict non-experimentally determined changes in transcripts, proteins, and metabolites due to a perturbation, usually gleaning information from a low-dimensional manifold that exhibits some biological meaning (e.g., a gene regulatory network – GRN) ***(sub-objective O1A)***. Other models in this category try to *elucidate* phenotypic characteristics of cells under perturbation pressure, usually by removing confounding sources of variation. These characteristics can reflect novel cell states following perturbed differentiation trajectories (across time when possible) or sometimes can allude to drug sensitivity estimates (e.g., IC50 values) ***(sub-objective O1B)***. A second objective relates to ***predicting the mode of action of a possible perturbation*** **(O2)**. For example, classification approaches can be deployed using the transcriptional changes a compound elicits on a cell type to predict whether this compound could be beneficial in reversing disease-bound cell phenotypes, thus facilitating novel drug repurposing or repositioning paradigm. Another objective is the capacity to predict ***synergistic or antagonistic interactions*** **(O3)** among various perturbations when envisioning combinatorial treatments. The perturbations can be genetic or chemical, and the analysis can be based on categorical (e.g., name of compounds) or continuous parameters (e.g., synergy scores); hence, various classification and regression approaches can be tested. The fourth objective stated by Ji et al. was to predict ***chemical structural components from biological perturbational data*** **(O4)**. A paradigmatic case of this objective is *hit identification* from gene expression data using artificial intelligence (AI), e.g. DrugMonizome [5] for bulk omics. By leveraging drug-drug similarity metrics and drug-gene information, ML models can propose novel chemical molecules that could propagate the desired -omic signature.

3. Which single-cell technologies and datasets can be used for perturbation modelling?

Understanding the various single-cell technologies and datasets that can be used for perturbation modelling is imperative to accomplish some or all of the aforementioned objectives.

Many perturbation studies revolve around single-cell datasets with one or more omic modalities arising from typical immunophenotyping of a specific tissue/organ/cell population or comparative studies dissecting normal and diseased cellular populations. On the other hand, there is also a burgeoning number of ***single-cell perturbational screens***, which we will briefly address in this section. In this section, we concisely introduce the fundamental single-cell technologies underpinning this field, and we encourage the readers to explore an excellent review by Cheng et al. [6], which delves more in-depth into this particular subject.

Some initial groundbreaking technologies in this field include **Perturb-seq** [7] **and CRISP-seq** [8], which are highly parallelised single-cell technologies built on CRISPR-pooled screens. These techniques trace the identity of single guild RNA (sgRNA) by incorporating *unique guide barcodes (GBC)* into sgRNA-encoding plasmids. The endogenous mRNA and GBC transcripts are ligated by the same cell barcode (CBC) during reverse transcription (RT) within the built-in microchambers of single-cell platforms. Two barcodes within single sequencing read transcriptomes, and corresponding perturbations can be associated via mapping. Notwithstanding, the aforementioned technologies are compounded by pitfalls like the template-switching effect during viral co-packaging, which remain challenging unless the homology transfer vector or an arrayed package is utilised. **CROP-seq** [9] (polyadenylated sgRNAs) has the critical advantage of reading the guide RNA directly, which streamlines single-cell CRISPR screening with extensive guide RNA libraries, making CROP-seq fully compatible with standard cloning protocols for pooled CRISPR screens. However, CROP-seq also faces particular challenges, such as the excessive length of this nested cassette, which prevents multiplexed perturbation and decreases the rate of sgRNA retrieval. To tackle this challenge, Replogle et al. pioneered the **direct Perturb-seq approach**. This method leverages a capture sequence (CS) integrated at selectable locations within the sgRNA construct and complementary primers designed to capture sgRNA transcripts directly. **Direct-seq** is a variation of this latter approach involving an A/G mixed capture sequence added to the gRNA scaffold [10].

MIX-seq [11] is another noteworthy technology that enables the study of perturbation responses to compounds or genetic interventions from various cell groups and the disentanglement of cell identity based on single-nucleotide polymorphism (SNP) profiles. **sci-Plex** [12], on the other hand, is a single-cell combinatorial indexing approach for measurement of the global transcriptional responses to thousands of different perturbations using "nuclear hashing". Moreover, **Tap-seq** [13] is a sensitive, low-cost, and platform-independent technique targeting specific genes for scRNA-seq coverage. **PoKI-seq** [14], combines single-cell transcriptome analysis and pooled knockin screening.

Another niche within these technologies is single-cell perturbation datasets with multiple omic modalities (e.g., proteins). **ECCITE-seq** [15] and **Perturb-CITE-seq** [16] are two paradigmatic cases that entail in tandem detection of transcripts, proteins, and sgRNAs. Of note, **Mosaic-seq** [17] is a noticeable technique that involves an epigenetic and transcriptomic mosaic in affected cells due to successive infections with vectors promoting the expression of the KRAB epigenetic repressor and a library of different sgRNAs, ultimately leading to targeted epigenetic changes and altered gene expression. **sc-Tiling** [18] is introduced as a technique to combine single-cell transcriptome and protein structure analysis with a CRISPR gene-tiling screen. Additionally, **Perturb-ATAC** [19] combines genome-wide chromatin accessibility profiling in single cells based on the simultaneous detection of CRISPR guide RNAs and open chromatin sites by assaying transposase-accessible chromatin with sequencing (ATAC-seq) and multiplexed CRISPR interference or knockout (KO). The **Spear-ATAC** [20] technique combines integrated sgRNA spacer sequences and chromatin accessibility profiles from hundreds of individual cells. A similar approach involves CRISPR screens with **CRISPR–sciATAC** [21], which enables the joint capture of chromatin accessibility profiles and CRISPR perturbations. Compared to Perturb-ATAC, both Spear-ATAC and CRISPR–sciATAC offer higher throughput. For example, CRISPR-sciATAC uses two-step combinatorial indexing to label DNA molecules with unique cell barcodes and requires no specialised equipment. Furthermore, CRISPR–sciATAC can generate thousands of single cells at ~20 × less reagent cost and requires ~14 × less time than Perturb-ATAC. The following table references an overview of paradigmatic cases from the various single-cell perturbation datasets with their cellular context. Table 1.

We also observe the onset of integrated databases such as **scPerturb**, which harmonises 44 publicly available datasets (http://projects.sanderlab.org/scperturb/) [26]. This unprecedented effort for dataset

Table 1
Indicative examples of perturbation modelling datasets.

Technology	Omic modality	Indicatively produced datasets
Perturb-seq	RNA	*lung cancer cells (200 TP53 and KRAS variants in 300.000 cells)*[22], *MOLM13 AML cells (28 mSWI/SNF targeted genes)*[23], *MC38 cells (3 targets:Prmt1, Ripk1, Axl)*[24], *Sars-CoV-2 infected Calu-3 cells (183 targets)*[25]
CROP-seq	RNA	*Jurkat cells, 116 sgRNAs (32 genes)*[9]
Direct-seq	RNA	*mKate2 breast cancer cell line, 12472 sgRNA pairs*[10]
MIX-seq	RNA	*24-99 cell lines, 1-13 perturbations, 1-5 timepoints, 4 small molecule experiments and 1 genetic experiment*[11]
Perturb-CITE-seq	RNA/protein	*Patient-derived melanoma cells and autologous tumor-infiltrating lymphocyte (TIL) co-cultures, 20 proteins, ~218.000 cells, ~750 perturbations for targets related to resistance against immune checkpoint inhibitors*[16]
sci-Plex	RNA	*188 perturbations, A549-K562-MCF7 cell lines, 4 dosages, 2 timepoints*[12]
Tap-seq	RNA	*1778 enhancers in 2 chromosomes of K562*[13]
PoKI-seq	RNA	*36-member library of targets, Human T cells*[14]
sc-Tiling	RNA	*602 sgRNAs targeting the coding exons of mouse Dot1l in MLL-r leukemic cells*[18]
Perturb-ATAC	DNA	*40 sgRNA genotypes in 2627 single immortalized B lymphoblasts*[19]
Spear-ATAC	DNA	*~20 targets, various time-points in K562 cells*[20]
ECCITE-seq	RNA	*infected K562 cells with a CRISPR library comprising guides targeting genes encod- ing cell surface markers (CD29 and CD46), intracellular signaling molecules (JAK1 and p53)*[15]
Mosaiq-seq	RNA/ epigenome	*K562 cells, 241 sgRNAs spanning 71 constituent enhancers from 15 super-enhancers within seven topologically associated domains (TADs)*[17]
CRISPR-sciATAC	epigenome	*K562, 21 chromatin modifiers*[21]

harmonisation involves quantifying perturbation strength and comparison of experiment-specific variables, such as the number of perturbations and the number of cells per perturbation, using the E-distance metric. E-distance represents a statistical measure of the distance between two distributions, which offers an intuition about the signal-to-noise ratio in a dataset. A low E-distance indicates that a perturbation did not induce a significant shift in expression profiles, reflecting technical problems in the experiment, weak effect of the perturbation, or resistance to perturbation. Across datasets, E-distances between perturbed and unperturbed cells varied notably. The NormanWeissman2019 dataset stood out with the largest mean E-distance compared to similar-sized datasets, indicating significant differences in expression profiles between perturbed and unperturbed cells. This dataset's distinctiveness likely originates from using two-target perturbations with CRISPRa, where targeting the same gene with two single guides increased the chances of causing considerable transcript profile changes. scPerturb offers all 44 datasets in the form of.h5ad format.

Lastly, the PertPy Python package (https://pertpy.readthedocs.io/en/latest/installation.html) is a well-documented framework under active development that contains several perturbation datasets in.h5ad format, an assortment of perturbation modelling tools, and tutorials for deployment.

4. Perturbation models

In this section, we will describe several *in silico* methods for perturbation modelling spanning from various biostatistics models to more ML-oriented architectures and models inspired by GRNs. We provide a coarse description of each method's algorithmic and mathematical underpinnings to enable the reader to contextualise these computational tools succinctly.

To group the perturbation tools in a tractable way for the general audience, we constructed a matrix in a Boolean logic (0 =False, 1 = True) with perturbation features (*OOD, Perturbation MoA, Perturbed cell states, Predict combinatorial perturbations, Pharmacogenomic latent space/ Drug Repurposing, Removing confounding effects, Digital KOs, Causal inference*), computational features inspired by the approach by Ji et al., 2021 (*Shallow model, Matrix factorisation, Autoencoder_based, GRN-based, Distance metrics, Foundational models, Hybrid models_no_DL, Hybrid models_DL*) and finally dataset-specific features (*Time-series, spatial*). By running the UMAP algorithm for dimensionality reduction, subsequent K-means clustering and applying domain knowledge, we could pinpoint six groups of perturbation modelling tools with evident computational and biological coherence (Supplementary Table 1).

In Fig. 1., we provide a general overview of the perturbation methods clustered by their functionality, and in Fig. 2., we give some representative examples of perturbation modelling methodologies in a more fine-grained manner.

We've developed a supplemental document featuring definitions of key concepts to enhance comprehension of the intricate statistical and computational terms associated with the tools described here. This resource aims to assist readers with a more experimental biological focus on Supplementary Table 2**.** All listed tools have active GitHub pages, as shown in Supplementary Table 3.

4.1. Shallow models

This first group of tools entails several classical statistical inference models, which refer to statistical methods and techniques typically used to conclude populations based on sample data. These models are predicated on the principles of frequentist statistics and rely on assumptions such as random sampling and a well-defined probability distribution. Furthermore, this group also includes specific shallow, interpretable ML tools (classifiers, regressors) that do not contain a highly perplexing computational architecture.

MUSIC [27] is an analytical tool for single-cell CRISPR screening data analysis. By allocating *topic probability profiles* instead of discrete clusters and enabling a quantitative assessment of perturbation impact, MUSIC automatically chooses the ideal topic number for perturbation analysis and handles problems such as sample imbalance. At its core, MUSIC considers a single cell with perturbation as a document and the gene expression as the word frequency in the document.

scMAGeCK [28] detects genes, including non-coding elements, linked to various phenotypes in single-cell CRISPR screening data. It is an improved version of MAGeCK that utilises only pooled CRISPR screens and uses scRNA-seq data as the screening experiment's readout. The scMAGeCK framework is composed of two components: (i) scMAGeCK-Robust Rank Aggregation (RRA) which is a sensitive and precise algorithm for detecting genes linked to a single marker expression, and (b) scMAGeCK-LR, which is a linear-regression-based approach that reveals perturbation effects on thousands of gene expressions.

Drawing from the paradigm of negative binomial distribution in single-cell data [29], **SCEPTRE** [30], analyses single-cell CRISPR screens using conditional resampling to tackle batch effects and diverse sequencing depth, which are confounding sources of variation masking actual perturbational effects. Built on the conditional randomisation test, SCEPTRE provides robust statistical analysis, effectively removing confounding noise. The method uncovers numerous *cis* and *trans--regulatory* relationships and proves valuable in dissecting regulatory mechanisms related to GWAS associations.

Augur [31] prioritises cell types based on their molecular response to perturbations in highly multidimensional single-cell data. The approach aims to quantify the separability of cells undergoing a significant response compared to those unaffected, considering heteroscedasticity and variability within cellular subpopulations. Augur achieves this by training cell-type-specific machine-learning models (random-forest classifiers) to predict experimental conditions (e.g.,

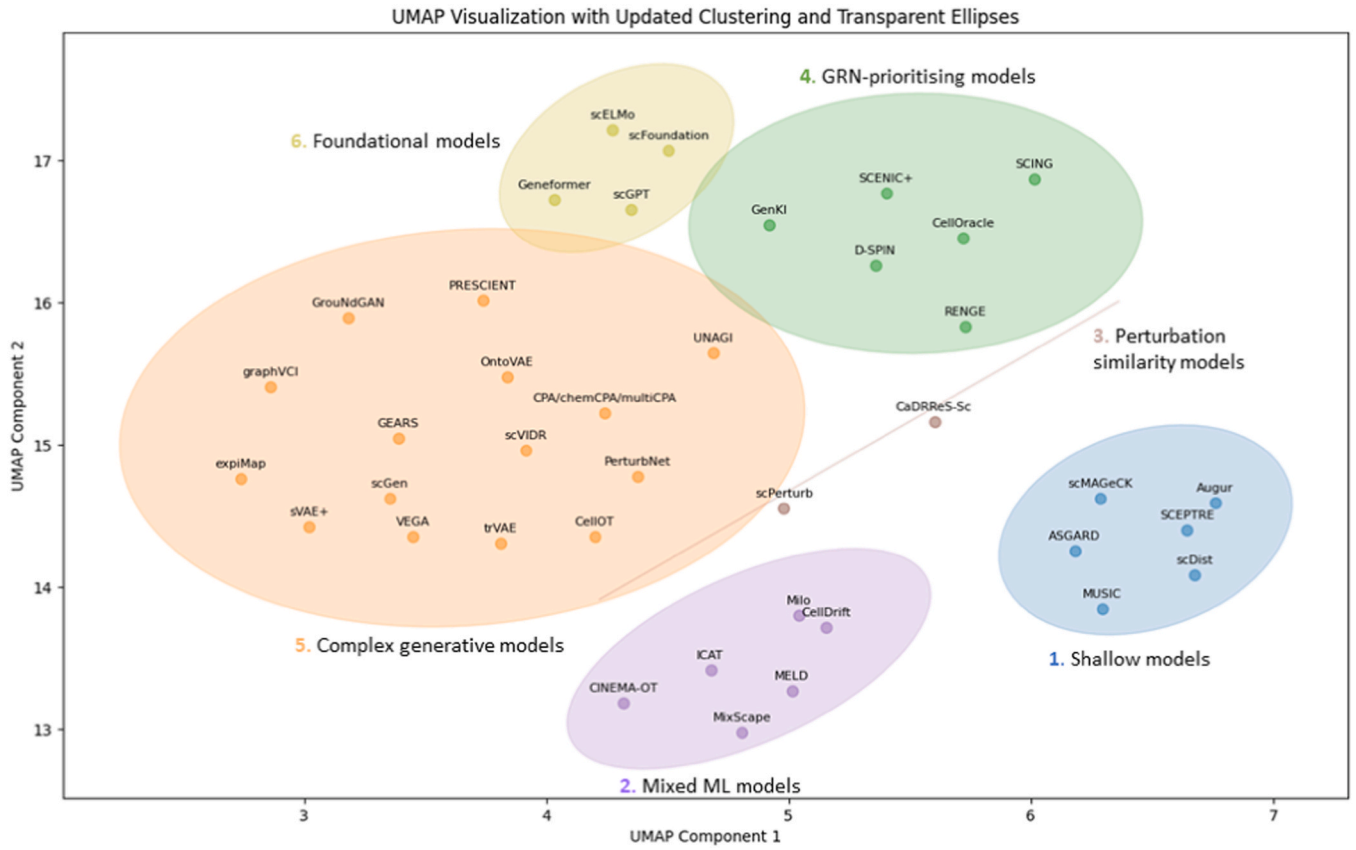

Fig. 1. Computational tools for perturbation modelling, clustered by their functionality. A UMAP plot of the 41 perturbation tools reviewed in the present manuscript, clustered into 6 distinct clusters by k-Means clustering. Supplementary Table 1 provides each tool's perturbational, computational, and dataset-specific features used to perform the clustering.

treatment versus control) solely from molecular measurements. The method is versatile, accommodating various molecular feature modalities such as scRNA-seq, proteomics, epigenomics, and imaging transcriptomics datasets. Augur employs a Random Forest classifier, known for its non-parametric nature, making it robust to different molecular measurements and pre-processing steps, surpassing the Wilcoxon test usually used for differential expression analysis in these scenarios.

Inspired by the Principal component regression analysis, **scDist** [32] explicitly models individual and technical variability, providing a quantitative definition of group transcriptomic differences. scDist relies on a linear mixed-effects model for single-cell gene expression counts. Furthermore, given the high-dimensional nature of transcriptomic profiles, scDist utilises an approximation for between-group differences through a low-dimensional embedding, thus quantifying the statistical uncertainty due to individual-to-individual variation and other sources of variability. scDist performs robustly in small patient cohorts (*a mainstay in most single-cell studies*) where it is always challenging to extract statistically robust conclusions.

Based on differential expression tests and order statistics, **ASGARD** [33] performs drug repurposing by creating a dependable drug score for repurposing drugs by using consistently differentially expressed genes as inputs to identify drugs that can significantly reverse their expression levels based on the L1000 drug response dataset. To designate drugs for multiple clusters, ASGARD defines a drug score to evaluate the drug efficacy across multiple cell clusters. The drug score considers various factors such as the proportion of cell types, the significance of reversing the differential gene expression pattern for each selected cell cluster, and the proportion of significantly deregulated genes that the drug treatment can reverse in each selected cell cluster.

4.2. *Mixed ML models*

This perturbation model group contains hybrid approaches that blend distinct biostatistical and ML approaches, such as GLMs, dimensionality reduction, clustering, and probabilistic learning.

Milo [34] is a method to detect compositional changes occurring in smaller subpopulations of cells, designated as neighbourhoods over the k-nearest neighbour (KNN) graph of cell-cell similarities. In Milo, per-neighbourhood cell counts are modelled using a negative binomial generalised linear model (GLM), and hypothesis testing is conducted to retrieve differentially abundant neighbourhoods. By representing cell states as overlapping neighbourhoods, Milo accurately identifies emerging perturbed cellular states (without the need of a priori cell annotation), facilitating the identification of underlying disease-related molecular motifs.

Mixscape [35] unmasks actual perturbation effects from confounding sources of variation, such as cells escaping perturbations. Mixscape is based on mixture discriminant analysis (MDA), which assumes that individual samples fall into different groups but that each group is a mixture of distinct subclasses. This is particularly relevant to data from CRISPR experiments, in which individual cells can be grouped based on their expressed gRNA. Still, each group can represent a mixture of 'perturbed' and 'escaping' (or non-perturbed) subclasses since the perturbation effect in a cell group is not strictly binary (yes-no) in real-life cell biology. MDA facilitates the interpretation of each target gene class as a mixture of two Gaussian distributions, one representing the knockout (KO) and the other the non-perturbed (NP) cells. Mixscape assumes that the distribution of the NP cells is identical to that of cells expressing non-targeting gRNAs (NT) and hence tries to gauge the distribution of KO cells. A paradigmatic implementation of Mixscape

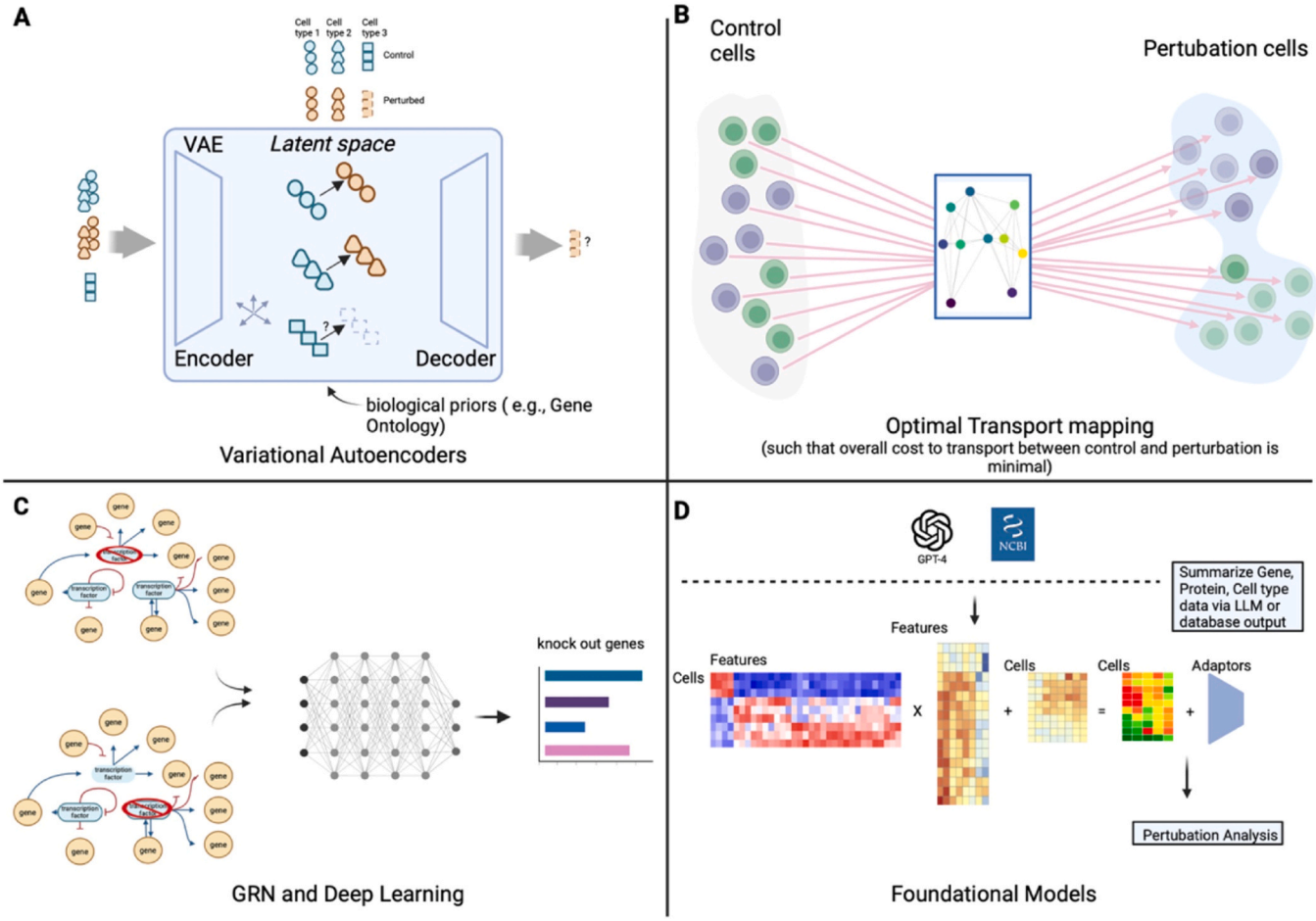

Fig. 2. Snapshots of perturbation modelling concepts in single-cell biology. (A) Variational Autoencoders (VAEs) predict unseen perturbations absent from the respective training data for latent space learning. Some of these approaches encompass biological priors (e.g., OntoVAE - Gene Ontology) to enable more biologically plausible predictions. (B) OT mapping concept showcasing how this mathematical theorem can be used to predict the distributions of perturbed cells from unpaired control measurements. (C) Gene Regulatory Network (GRN)- inspired models provide opportunities for digital knock-outs of potentially critical transcription factors, utilising the non-linear extrapolation of neural networks to discern which genes become more affected. (D) The latest trend in single-cell perturbation modelling draws inspiration from large foundational models combining embeddings from LLM models, and bioinformatic databases with single-cell omics to formulate perturbational predictions.

entails ECCITE-seq data, in which it substantially augments the signal-to-noise ratio for downstream analyses.

MELD-VFC [36] regards single-cell experiments as samples from a probability density function over the underlying transcriptomic manifold plane. MELD-VFC quantifies the perturbation effect by calculating alterations in the probability density of the experimental condition compared to the control. Then, MELD-VFC calculates sample-associated relative likelihoods to discern whether a cell belongs to the control or the perturbed population. On this basis, the frequency composition of sample labels and relative likelihood scores form the basis for the **Vertex Frequency Clustering (VFC) algorithm**. Moving away from unsupervised clustering approaches based on mere data geometry, *VFC clusters cells that exhibit similar responses to a perturbation* (enriched, depleted, or unchanged), providing a detailed view of perturbation responses. Lastly, the differential expression between VFCs allows the extraction of gene signatures associated with a perturbation.

CellDrift [37] builds upon the implementation of GLM on Functional Data Analysis (FDA) to explore temporal cell group profiles shaped by longitudinal perturbations. Functional data analysis (FDA) enables the study of multiple curves changing over time, where each curve is a sample tracing with a series of time points, which can be characterised as a function (such data are called functional data). CellDrift's GLM recovers temporally perturbed genes, and the GLM contrast coefficients are provided as an input to the Fuzzy C-mean algorithm to uncover temporal patterns. Then, functional principal component analysis (FPCA) identifies the dominant modes of temporal variance in functional data. Ultimately, CellDrift can pinpoint temporal trends in genes and biological pathways across distinct cell groups.

ICAT [38] is a two-test algorithm (sparse feature weighting with semi-supervised Louvain community detection) *for robustly and sensitively identifying cell states* across treatments in scRNA-seq data. ICAT begins by performing self-supervised feature weighting to discern genes that distinguish cell identities among control cells, followed by semi-supervised clustering using the transformed expression matrix. ICAT clusters control cells with Louvain community detection to learn feature weights and then use these clusters as input into NCFS (Neighborhood Component Feature Selection) to weight genes based on predictiveness. After applying these gene weights to the original expression matrix, ICAT clusters treated and control cells together using a semisupervised Louvain community detection method, maintaining the previously identified cluster labels for control cells unchanged. Due to the sparse nature of NCFS, ICAT removes technical and biological noise by focusing on the clustering of highly informative genes and hence exhibits noticeable biological interpretability in discerning perturbed cell states.

A recent trend in the field is that the Optimal Transport (OT) concept provides a mathematical framework for analysing probability

distributions. OT has already been proposed for single-cell applications, such as integrating multiple single-cell datasets [39]. **CINEMA-OT** [40] is a framework for causal learning and optimal transport that aids in distinguishing between intrinsic cell-state effects (confounders) and perturbation effects for scRNA-seq datasets. CINEMA-OT utilises independent component analysis (ICA) and filtering, guided by a functional dependence statistic (*Chatterjee's coefficient-based distribution-free test*), to discern and separate confounding and perturbation-dependent factors. CINEMA-OT uses OT with entropic regularisation to generate causally matched counterfactual cell pairs. CINEMA-OT can be applied for individual treatment-effect analysis, response clustering, attribution analysis, and synergy analysis.

4.3. Perturbation similarity models

The following methods gravitate primarily toward capturing drug similarity from transcriptomic data and chemo-bioinformatic databases, which are leveraged for prior knowledge.

CaDRReS-Sc [41] (Cancer Drug Response prediction using a Recommender System for single-cell RNA-seq) is the single-cell version of the CaDRReS model (https://doi.org/10.1093/bioinformatics/bty452), which predicts cancer drug responses for unseen cell lines or patients, based on learning projections for drugs and cell lines within a ***latent 'pharmacogenomic' space.*** The model entails matrix factorisation and is trained on biological prior knowledge from the Genomics of Drug Sensitivity in Cancer (GDSC) database (1074 cancer cell lines and 226 drugs). CaDRReS-Sc can effectively combine cell-specific drug response values into a holistic response value using Newton's method, which is ideal for constructing IC50 sigmoid-like curves. CaDRReS-Sc combines cell- or cluster-specific drug response predictions to estimate overall drug response and prioritise drug combinations for a patient. Moreover, CaDRReS-Sc enables drug-pathway associations and combinatorial therapy responses for personalised therapeutic regimens.

Interestingly, the **scPerturb** [42] framework (referenced in the manuscript's perturbational data section) also uses the E-distance metric to reveal similarities between two or more perturbations, facilitating studies for potential synergistic treatments. This type of analysis was conducted on the Papalexi2021 and the NormanWeissman2019 Perturb-seq datasets, revealing that permutations in molecular targets that are second messages to the same pathway elicit paradoxically similar phenotypes (i.e. small E-distance).

4.4. GRN-prioritising models

Networks, or graphs, permeate biomedical research, from molecular interaction maps to dependencies between diseases in a patient to populations encompassing social and health interactions. Gene Regulatory Networks (GRNs), which try to illustrate the reciprocal interactions of genes and their targets (regulons), are an essential facet of biological networks and can be used in perturbational modelling as priors to facilitate more plausible predictions [4,43]. The following tools are mostly predicated on GRN analysis**.**

CellOracle [44,45], is an innovative tool that seamlessly integrates scRNA-seq and scATAC-seq datasets to pinpoint crucial transcription factors controlling gene targets within understandable network models. By leveraging gene-regulatory networks derived from these data types, CellOracle performs *in silico* transcription factors perturbations, effectively simulating cell state changes in response to TF perturbations. It combines promoter and enhancer peaks from scATAC-seq with scRNA-seq data to create cluster-specific GRN models. These models allow CellOracle to project the outcomes of TF perturbations onto cell trajectory maps. For instance, CellOracle can simulate cell identity transitions following digital (*in silico*) TF perturbations such as knockout or overexpression. It calculates shifts in target gene expression, estimates transition probabilities by comparing these shifts with local neighbours, and then translates these probabilities into weighted average vectors to indicate the direction of cell state transition for each impacted cell.

Notably, **D-SPIN** [46] identifies gene programs that are co-regulated groups of genes through unsupervised orthogonal nonnegative matrix factorisation (oNMF) and phenotypic gene-program annotation. D-SPIN constructs a unified regulatory network with edges representing predicted interaction between gene-expression programs and edges approximating interactions between the network and applied perturbations. D-SPIN uses different versions of maximum-likelihood inference to capture these interactions. The output of D-SPIN concerns the calculation of probabilities of transcriptional states across perturbations and perturbation grouping. In Perturb-seq data and single-cell data from compound treatments, D-SPIN effectively uncovers the organisation of cellular pathways the regulatory logic governing transcription/translation/metabolism and unravels molecular mechanisms behind synergistic pharmacological regiments.

RENGE [47] is a computational approach for deducing GRNs by analysing time-series single-cell CRISPR data. RENGE reconstructs a signed GRN from the time-series expression data after gene KO, obtained by scCRISPR analysis, by modelling how the effects of the gene KO propagate on the network. RENGE can differentiate between direct and indirect regulatory interactions, facilitating the inference of regulations by genes not subject to knockout. RENGE's resultant network aligns well with multiple databases and existing literature.

Moreover, **SCENIC+** [48], an improved version of SCENIC [49], utilises chromatin accessibility through 30.000 + known motifs and gene expression profiles at the single-cell level and offers a chance to infer enhancer-driven gene regulatory networks (eGRNs). Notably, SCENIC+ anticipates genomic enhancers and identifies potential upstream transcription factors (TFs) and establishes connections between these enhancers and their relevant target genes. This approach employs a Random Forest regression model for feature selection, fitting individual models for each gene by utilising TF expressions as predictors for gene expression. Then, the TF expression is manipulated *in silico*, and the resulting impact on gene expression is forecasted using these regression models. This iterative process is repeated to simulate indirect effects. The original and simulated gene expression matrices are jointly embedded in the same dimensionality reduction, allowing for the visualisation of the anticipated effects of the perturbation on cell states.

SCING [50] is a novel GRN inference method. Combining supercell designation, gene neighbourhood-based connection pruning, bagging, gradient boosting regression, and conditional mutual information approaches to identify robust regulatory relationships between genes based on single-cell transcriptomics data, SCING demonstrates robustness in identifying GRNs from diverse datasets, including scRNA-seq, snRNA-seq, and spatial transcriptomics. Performance evaluations using Perturb-seq datasets, held-out data, the mouse cell atlas, and the DisGeNET database reveal that SCING surpasses existing methods in accuracy and biological interpretability.

Another pertinent method is **GenKI** [51], which employs an unsupervised variational graph autoencoder (VGAE) model which learns the latent gene representations of wild-type (WT) samples and virtually constructs a virtual KO counterpart to discern similarity. GenKI initially creates a GRN model using a PC regression method and applies a thresholding technique to prune potentially spurious TF-gene connections. Then, the GRN is provided as input in the VGAE. GenKI virtually ablates gene expression by computationally removing all edges associated with the KO gene from the GRN model. Subsequently, distinctions between WT and virtual KO data are discerned by comparing their respective latent parameters obtained from the trained VGAE mode through the Kullback-Leibler (KL) divergence (higher KL divergence values indicate more significant impacts of the KO on specific genes). A bagging-based method is then used to identify genes significantly perturbed by the KO. Then, the enriched functions of these perturbed genes (KO-responsive genes) are analysed to predict the functions of the KO gene, leveraging similarities in affected pathways or biological

processes.

4.5. Complex generative models

Due to the vast number of cells in single-cell omics datasets, typically in the tens of thousands, Deep Learning (DL) and relevant generative models are being cardinally implemented in this field. Such datasets generate a voluminous dataspace that can be used to implement this non-linear distribution model.

One of the most well-researched tools in this category is **scGEN** [52], which leverages vector arithmetics in the latent space of VAEs, a technique that assumes the cellular response induced by stimuli can be modelled linearly. This latent space represents a low-dimensional manifold that arises by underlying constraints imposed by homeostatic cellular mechanisms like GRNs. The principle behind scGEN is that in the low-dimensional manifold, there is significant linearity that can capture variations elicited by the effect of a perturbation deriving from the high-dimensional single-cell dataset. Furthermore, with considerable accuracy, scGEN was robust in facilitating cross-study and cross-species perturbation predictions. scGEN is on of the most well-researched models in the field and is frequently used in small comparative studies.

An alternative model is **trVAE** [53], which incorporates maximum mean discrepancy (MMD) in a conditional variational autoencoder (CVAE) framework. The MMD regularisation forces representations in the first layer following the bottleneck to be similar across conditions. This regularisation helps prevent overfitting and promotes the learning of a more meaningful and transferable latent space. Contrary to scGEN, trVAE can address the issue of performing OOD predictions for multiple perturbations within a given dataset.

Compositional perturbational autoencoder (**CPA**) [54] uses an adversarial autoencoder framework to find embeddings in the latent space for drug perturbations, covariates, and basal cellular states in single-cell drug screening datasets (e.g., sci-Plex). An extension of CPA called **chemCPA** [54] introduces into the autoencoder architecture a perturbation network that encodes small molecules using their known chemical descriptors, thus enabling extrapolations to unseen drugs. A recent version of CPA called **MultiCPA** [55] combines perturbation modelling with multi-modal inferencing (i.e., leverage paired measurement of RNA and epitopes to approximate perturbation responses across single-cell modalities).

VEGA [56], (VAE enhanced by gene annotations), is a VAE featuring a sparse linear decoder guided by biological networks guided by gene module membership as recorded in gene annotation databases (e.g., Gene Ontology, PANTHER, MolSigDB, or Reactome). VEGA can aid the disentanglement of cell types and states while robustly performing out-of-distribution predictions. Concerning the latter, compared to scGEN, VEGA sacrifices predictive performance for biological interpretability of the latent space.

scVIDR [57] applies linear and log-linear interpolation on vector arithmetics to calculate differentially expressed genes in a cell-type-specific fashion (*contrary to scGEN, which calculates vector arithmetics in the latent space in a non-cell-type specific manner*) and extrapolate perturbation effects to unseen cell types and drug dosages, respectively. To approximate the function of the decoder, scVIDR uses ridge regression, which provides an explainability "flavour" to the latent space of the VAE architecture by generating "gene scores". These scores represent how significant changes in latent space dimensions will impact the decoded transcriptomic response (i.e., genes with high gene scores will be heavily affected by the predicted perturbation intervention). scVIDR also introduces the concept of "pseudo-dose" trajectory. Considering that cells of the same type have variable sensitivities to the same perturbation, the "pseudo-dose" trajectory provides a scalar co-efficient of perturbation effect for each singular cell, enabling a better representation of the underlying perturbational biology.

sVAE+ [58] is a variant of sVAE model with a Bayesian approach for learning sparse mechanism shifts that require minimal hyperparameter tuning. This model builds upon the concepts of *disentangled representations and causal inference*, aspiring to designate representations of perturbations and cells that are more mechanistically explainable (i.e., how does a GRN change due to a perturbation?) and more efficient for out-of-domain generalisations. **sVAE+** tries to explicitly model cellular perturbations as interventions on latent variables to understand causal mechanisms, while incorporating a model of experimental noise from single-cell RNA-seq assays.

Recently, **graphVCI** [59] emerged as a new approach combining deep graph representation learning with variational Bayesian causal inference to predict single-cell perturbation effects, approximating causal machine learning. graphVCI entails an adjacency matrix updating technique producing refined relation graphs before model training that was able to discover more relevant relations to the data.

A vision for this field of single-cell perturbation biology is to create "perturbation atlases" of limited single-cell datasets, which could help to understand disease-driven or drug-induced changes. These would expand on previous projects and facilitate analysis, allowing for ML predictions [4]. A recent pertinent method is **expiMap (explainable programmable mapper)** [60]. This interpretable conditional variational autoencoder incorporates domain knowledge in the form of Gene Programs (GPs) into DL architectures for reference mapping of query disease perturbation. expiMap designates biological processes in normal and disease states when mapping the new queries to the atlas while maintaining comparable integration performance to relevant state-of-the-art tools. expiMAP accomplishes the above by performing "architecture programming," i.e., placing constraints through attention mechanisms to ensure that latent representations capture the variability of relevant GPs. expiMap can be more specific in retrieving biological pathways (in the form of GPs) that emerge as deregulated in a perturbation query dataset compared to the healthy reference atlas, outperforming traditional GSEA, which is prone to more generic enrichments. expiMap can also learn new GPs not previously included in the reference atlas specific to the perturbed query dataset. At the same time, it can approximate diverse treatment responses for patient cohorts and deconvolute cellular heterogeneity of intricate samples like tumour-microenvironment single-cell data.

CellOT [61] incrementally alters molecular profiles of cells (gene expression, cell signalling), which can be modelled using OT. To capture cell-state transitions upon a perturbation, CellOT generates an optimal transport map for each perturbation in a fully parameterised and highly scalable manner via input convex neural networks. CellOT tries to find a function (Tk) that ensures that the overall transport cost between the control and perturbed states is minimal. When applied to the latent representations from *state-of-the-art* autoencoders like scGEN, CellOT outperforms the baseline models by capturing better the sometimes subtle and nuanced cellular heterogeneity under the influence of a perturbation. In single-cell datasets containing drug treatments, CellOT excels in predicting cell-to-cell variability in drug responses, drugs' mode-of-action (MoA), plausible OOD extrapolations for cells and patients, reconstructing innate immune responses across species and in disentangling subpopulation-specific drug effects.

Also, **PRESCIENT** [62] is a generative modelling framework based on diffusion maps designed to fit longitudinal single-cell datasets. PRESCIENT incorporates a drift term (guided by a potential function) and noise term, with a neural network parameterising the drift function based on PCA-transformed gene expression data. Stochastic simulation predicts population states at future time steps, and model parameters are optimised to minimise the discrepancy between simulated and observed data points. In contrast to other approaches, like scGen, PRESCIENT models decipher time-driven cellular changes, generate cell distributions, and explicitly model cellular differentiation. Furthermore, PRESCIENT is apt for reconstructing potential cell fate trajectories after simulating digital KOs (singular or combinatorial) of crucial TFs.

GEARS [63] (graph-enhanced gene activation and repression

simulator) uses biological priors in the form of knowledge graphs and DL algorithms to predict unseen cellular responses to genetic perturbations, including single and multi-gene perturbations (perturbation set). Tested in Perturb-seq data, GEARS uses a gene co-expression knowledge graph as a prior when learning gene embeddings and a Gene Ontology (GO) knowledge graph when learning gene perturbation embeddings. A Graph Neural Network (GNN) combines information between neighbours in each graph. After cross-gene combination by a Multilayer Perceptron (MLP), the output is later fed in gene-specific output layers to enable *in silico* post-perturbation predictions for gene expression. Overall, GEARS predicts additive and non-additive combinatorial perturbation effects and can foreshadow new meaningful phenotypes arising from these perturbations.

PerturbNet [64] employs two distinct neural networks trained independently to encode extensive chemical or genetic perturbations and cell profiles into latent spaces. The output of these networks is provided as input to a third network, a conditional invertible neural network (cINN), which maps points from a perturbation space to a cell-state space. Subsequently, PerturbNet can predict gene expression changes caused by an unseen perturbation by encoding the perturbation, passing its representation through the cINN, and decoding the resulting cell states. PerturbNet can predict responses to coding sequence mutations apart from CRISPR and chemical perturbations and can attribute cell state shifts to specific features of perturbations like atoms from a compound's molecular structure. Also, PerturbNet can design perturbations to approximate a desired cell state distribution by minimising the square of the Wasserstein (W2) distance between counterfactual (i.*e., a theoretical distribution of control cells after a potential perturbation*) and the actual target distribution. PerturbNet is highly modular, allowing various architectures to be used in a data-specific manner (e.g., ChemicalVAE for compounds, GenotypeVAE for genetic interventions, and Evolutionary scale modelling network for amino acid sequences).

OntoVAE [65], a VAE with a multi-layer, sparse decoder, provides a different interpretable take on autoencoders by incorporating any hierarchical biological information encoded as an ontology (e.g., Gene Ontology, Human Phenotype Ontology). It confers direct interpretability in its latent space and decoder, allowing thousands of terms to be efficiently monitored without needing to preselect specific processes. Additionally, OntoVAE performs perturbation modelling while observing subsequent alterations in the activation of hidden nodes representing biological processes or phenotypes through the Gene Ontology schema.

GRouNdGAN (GRN-guided simulation of single-cell RNA-seq data using causal generative adversarial networks) [66] is a causally guided implicit generative model designed to simulate scRNA-seq data and *in-silico* perturbation experiments and evaluate GRN inference methods. Its architecture builds on the causal generative adversarial network and includes a causal controller, target generators, a critic, a labeller and an anti-labeler. The critic's role is to quantify the Wasserstein distance between the reference and simulated data. The target generators are trained in an adversarial manner to generate realistic data points indistinguishable from the reference data points by the critic. By incorporating a user-defined GRN into its structure, GRouNdGAN generates single-cell datasets reflecting both steady-state and transient-state conditions, wherein genes are expressed causally under the influence of their respective regulating transcription factors (TFs).

UNAGI [67] is a variational autoencoder with a generative adversarial network (VAE-GAN) that analyses time-series single-cell transcriptomic data. UNAGI generates a graph that captures temporal GRN motifs. Then, UNAGI performs perturbation modelling, either directly modulating drug target gene expressions or altering gene expressions by traversing the previously assembled GRN. Lastly, UNAGI facilitates drug repurposing by manipulating the latent space informed by real drug perturbation data from the CMAP database.

4.6. Foundational models

Generative pre-trained models, particularly those combining large and diverse datasets with pre-trained transformers, have excelled in natural language processing and computer vision. Inspired by an analogy between texts and cells (made up of words and genes, respectively), there is a growing interest in applying foundation models to cell and gene biology. For perturbation modelling, the aspiration is to illuminate the almost infinite combinatorial space of potential gene perturbations. These models can leverage the knowledge gained from cellular responses in known experiments and extrapolate to predict responses in unknown scenarios.

Geneformer [68] employs transfer learning, a technique within DL models, to establish a foundational comprehension of network dynamics. This is achieved through pretraining on around 30 million human single-cell transcriptomes that span diverse tissues and cell types found in public databases. The acquired knowledge is subsequently utilised for fine-tuning in various downstream applications with limited task-specific data, facilitating predictions in diverse settings. Geneformer accomplishes various tasks, including batch integration and cell-type annotation, predicting genomic elements, forecasting core networks and downstream targets, and employing *in silico* perturbation to predict phenotypes and model disease shifts to uncover potential therapeutic targets.

scGPT [69] is a novel single-cell foundation model pre-trained on over 10 million cells. With the help of self-attention mechanisms over the gene dimension, scGPT can encode complex interactions between perturbed and other genes. This enables the model to learn from existing experimental data (i.e., Perturb-seq data) through few-shot learning and make accurate predictions of gene expressions following perturbation for unseen experiments (OOD).

scELMo [70], leverages the benefits of Large Language Models (LLMs) and dedicated databases like NCBI to systematically establish a foundational model for analysing single-cell data. scELMo is predicated on transferring each cell's information from the sequencing data space to the LLM-embedded space. scELMo can predict potential "druggable" targets for a condition of interest by observing the change of embeddings corresponding to the removal of specific genes. Furthermore, scELMo can provide cell and feature embeddings in task-specific models (e.g., causal factor analysis in CINEMA-OT, gene expression predictions for cell-level and gene-level perturbations with CPA and GEARS, respectively) to simulate the effects of perturbations, outperforming the original iterations of these models.

Finally, **scFoundation** [71], is trained on a vast dataset of over 19, 264 genes with 100 million parameters on over 50 million scRNA-seq data, and it stands out as the most significant model regarding trainable parameters, gene dimensionality, and the number of cells used in pre-training. scFoundation, combined with DeepCDR and SCAD models, increases their baseline performance in IC50 inference and single-cell drug sensitivity prediction since it can robustly transfer cellular embeddings among bulk and single-cell data. By the same token, scFoundation was able to "up the ante" in perturbation modelling using the GEARS model by enabling cell-specific gene coexpression graphs, which the baseline model does not offer.

5. Small-scale comparative studies towards a unified benchmarking paradigm

Albeit the necessary small-scale comparative studies that accompany every new perturbational tool in its respective publication, there is a noticeable gap in the literature regarding systemic benchmarking with standardised pipelines, well-defined metrics, and specific biological endpoints for a holistic computational and biological evaluation. Most tools are compared with a few perturbational models on very few metrics and datasets. This lack of perturbational "best practices" represents one of the most significant impediments to a broader understanding and

implementation of these tools across diverse single-cell technologies.

Current small-scale computational evaluations in perturbation modelling entail metrics regarding correlations of differentially expressed genes (e.g., R2), distances between distributions (e.g., maximum mean discrepancy/MMD), clustering (Average silhouette width, Normalized Mutual Information, and Adjusted Rand Index) and machine learning performance (e.g., AUC, F1-score) [64,70] (Supplementary Table 4).

Acknowledging the small scope of these comparative studies, we will cite some noticeable examples to motivate further experimentation with some of these tools. A complete list of small comparative studies for each reviewed tool in this paper can be found in Supplementary Table 5. Concerning the GRN models, CellOracle surpassed other GRN-inference algorithms such as PIDC, GENIE3, and GRNBoost2 in accurately identifying biological ground truth [44], whereas SCENIC+ demonstrated superior performance compared to its original version, as well as ppcor, GRNBOOST2, PIDC, GRaNIE, FigR, and Pando, as evidenced by the AUC and F1 score metrics [48].

Regarding the complex generative models, the standout computational tool is scVIDR, which not only exceeded the performance of the traditional scGEN but also outperformed CellOT and scPreGan based on R2 metrics [57]. Meanwhile, OntoVAE shows comparable performance to VEGA and expiMap in terms of ARI metric [65], and sVAE+ surpassed both β-VAE and the standard VAE when comparing Pearson MCC scores [72].

6. Summary and outlook

Developing new treatments to address unmet therapeutic needs is challenging and resource-intensive, often with high attrition rates. Understanding the full impact of drugs on cells is difficult due to immense parameters shifting in cell physiology and an vast perturbation space involving various compounds, and dosages across time. However, computational tools in single-cell omics offer promise making these challenges more tractable. This field of "perturbation modelling" sits at the intersection of multiple disciplines and has gained momentum with advancements in single-cell technologies like entailing CRISPR genetic manipulations (e.g., Perturb-seq) or hundreds of compound treatments. Perturbation modelling can be encapsulated as diverse strategies that discern drug modes of action or extrapolate to unseen, counterfactual perturbation predictions.

The overarching vision for single-cell perturbation modelling could be contextualised within the mechanism-aware and multi-modal AI framework in Precision Mechanism [73]. Intersecting *in silico* perturbations with various omic datasets (e.g., metabolomics, imaging) and clinical Real-World Evidence (RWE) data, would be instrumental in facilitating tailor-made treatments to individual patients based on their underlying molecular idiosyncrasies. However, significant issues need to be tackled, hampering the broader application of these tools and the currently limited translational significance.

Firstly, despite Perturb-seq/CRISPR-seq-like datasets (Perturb-CITE-seq) and some time-series data, there is still a considerable scarcity of suitable single-cell datasets of higher-order complexity (large-scale longitudinal, spatial, multi-omic). Current perturbation modelling primarily captures a static snapshot of dynamic cellular ecosystems, impeding the inference of stochastic or deterministic molecular events over time or space. We expect the field to adopt more complex perturbation frameworks like the Perturb-Map [74] concept (*multimodal phenotyping of CRISPR screens* in situ *by imaging and spatial transcriptomics*). Furthermore, efforts like The Lifetime Consortia [75], which has a focus on interceptive medicine at the single-cell resolution, can drive the upscale of single-cell perturbation modelling efforts by providing vast resources of disease-omics with computational pipelines and the capacity to deliver in vitro “ground truths” through patient-derived experimental disease models (e.g., organoids). In line with the above, “perturbation atlases” for large patient cohorts as a concept is expected to make strides in enabling the integration of single-cell omics with clinical data from Electronic Health Records (EHRs). This will incentivise the co-estimation of single-cell perturbation modelling with clinical covariates like germline genetics, sex-related differences in pharmacokinetics and adverse drug reactions, thus providing a more holistic view of disease dynamics. We can also envisage the application of single-cell perturbation modelling in Anti-Microbial Resistance (AMR) based on the recent advances in droplet-based technologies for bacterial single-cell RNA-seq (BacDrop) [76].

Regarding computational methods, the current landscape is transitioning from models performing classical statistical inference and latent space learning to much more sophisticated methods incorporating shallow or deep learning models, with various mathematical theorems and GRNs as biological priors. The latest trend in the field is shaped by the emergence of large foundational models, such as LLM, which are trained on millions of single cells to create world models. We anticipate a significant increase in scientific endeavours toward refining these models for perturbation modelling. However, it is essential to exercise caution as notable weaknesses have been identified, particularly in zero-shot learning for fundamental single-cell analytical tasks such as batch correction and cell type clustering [77]. Still in its infancy, this niche of foundation models in perturbation single-cell technology could benefit from new LLM architectures like the Hyena layer or alternative architectures like the Diffusion model that could enable multi-omic and multi-modal single-cell datasets [78]. Additionally, we expect the implementation of LLM frameworks like **Patchscopes** in single-cell perturbation modelling, which can generate human-like text for “translating” the information in their representations for humans. Patchscopes herald a new era in the interpretability of LLMs, which could be of profound importance in unravelling the true molecular repercussions of a perturbation across cell type and the assumptions the perturbation model uses to reach each conclusion [79].

Another critical issue is the interpretability of the models, which has been a focal point of many efforts, as outlined by the various iterations of VAEs that we are reviewing in this manuscript. One potential approach to achieve this goal is expanding on *causal machine learning*. This technique involves constructing a causal perturbation model consisting of a graph-based model encoding genes as causal variables and an interventional model representing the perturbation. To provide more biologically plausible insights to biomedical scientists, the recently proposed *cellSCM prototype* could be combined with causal kinetic models, causal representation learning (CRL), and integration of biological priors [80].

Highly complex single-cell multi-omic data, especially with the advent of spatial omics, necessitate more powerful ML/DL approaches than ever to glean perturbational knowledge (e.g., the recently presented IMPA pipeline) [81]. To increase the robustness Multi-Task Learning (MTL) uses the information in training data as an inductive bias to perform multiple tasks in parallel, and the learning of one task improves the understanding of others through shared representations [82]. However, MTL requires tasks to have some form of correlation or semantic relevance between them. Contextualising affinities among latent representation learning with other pivotal tasks in single-cell analysis like GRN inference, cell clustering, multi-modal inference could be a viable way to allow DL architectures to capture better the systemic changes that a perturbation can confer across diverse cell populations. Recent tools like UnitedNet [83], adopted in a perturbational landscape, could be highly informative for better future inferences of multi-modal Perturb-seq models.

Integrating perturbation modelling into drug development and repurposing presents a promising frontier with multiple avenues for growth and innovation. While specialised tools for bulk-omics and multi-omics analyses abound, the field is still catching up in single-cell data analysis, with spatial omics remaining an area of emerging interest. Techniques such as multi-CPA and expiMAP promise more profound insights into cellular responses, mainly when applied to specific

perturbation atlases. Implementing distance metrics on single-cell datasets containing drug compounds, such as sci-Plex, can aid in quantifying similarities between different perturbations, facilitating comparative analyses. Moreover, causal machine learning applied to perturb-seq type datasets could systematically advance drug repurposing efforts while identifying and mitigating adverse drug reactions through methods like CINEMA-OT on sci-Plex datasets. We also envision that foundational models will make strides in streamlining drug repurposing efforts from single-cell data. Based on the paradigm of scElmo and scFoundation interacting with the GEARS pipeline, we can expect that future foundational models will be able to retrieve embeddings and form a multi-faceted *pharmaco-multi-omic latent space* from perturbation single-cell/spatial datasets (e.g., the CPJUMP1 resource dataset encapsulating three million images and morphological profiles of cells treated with matched chemical and genetic perturbations) [84], single-cell atlases (e.g., *Pan-cancer blueprint, Human Tumour Atlas Network, SC chromatin accessibility atlas*), chemoinformatic drug safety data platforms (FAERS, WHO-UMC) and other pharmacological and biological knowledge graphs [85].

Lastly, a more general remark is that with this bevy of datasets and computational toolkits, single-cell perturbation modelling is in dire need of large-scale benchmarking efforts to establish coherent metrics for related tasks as well as standards of metadata and file formats (e.g., ArrayExpress, HCA) to tackle interoperability issues. Collaborating with pertinent stakeholders and actively participating in open science communities, such as the Single Cell Community of ELIXIR (https://elixir-europe.org/communities/single-cell-omics), holds great potential for addressing these challenges. It would enhance the scientific community's involvement in the fascinating and ever-evolving domain of perturbation modelling.

CRediT authorship contribution statement

Fotis Psomopoulos: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing. **Naveed Ishaque:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Investigation, Project administration. **George Gavriilidis:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing, Software. **Aspasia Orfanou:** Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. **Vasileios Vasileiou:** Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Project administration, Software.

Declaration of Competing Interest

The authors declare no conflict of interest whatsoever.

Acknowledgements

The authors express gratitude for the financial assistance received from the HORIZON-INFRA-2021-EOSC-01-04 project **Scilake** (https://scilake.eu/) and **ELIXIR** (https://elixir-europe.org/), the research infrastructure for life-science data. Additionally, the authors acknowledge the financial support provided by the Federal Ministry of Education and Research of Germany in the framework of **SAGE** (project number 031L0265).

Code availability

The script to recreate the UMAP visualisation after K-means clustering and manual curation of the perturbation modelling methods is found in https://github.com/BiodataAnalysisGroup/PertReview.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.04.058.

References

[1] Ji Y, Lotfollahi M, Wolf FA, Theis FJ. Machine learning for perturbational single-cell omics. Cell Syst 2021;12:522–37.
[2] Subramanian A, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. e17 Cell 2017;171:1437–52. e17.
[3] Ye C, et al. DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nat Commun 2018;9:4307.
[4] Ji Y, Lotfollahi M, Wolf FA, Theis FJ. Machine learning for perturbational single-cell omics. Cell Syst 2021;12:522–37.
[5] Kropiwnicki E, et al. Drugmonizome and Drugmonizome-ML: integration and abstraction of small molecule attributes for drug enrichment analysis and machine learning. Database 2021;2021:baab017.
[6] Cheng J, et al. Massively parallel CRISPR-based genetic perturbation screening at single-cell resolution. Adv Sci 2023;10:2204484.
[7] Dixit A, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell 2016;167:1853–1866.e17.
[8] Jaitin DA, et al. Dissecting immune circuits by linking crispr-pooled screens with single-cell RNA-seq. Cell 2016;167:1883–1896.e15.
[9] Datlinger P, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods 2017;14:297–301.
[10] Song Q, et al. Direct-seq: programmed gRNA scaffold for streamlined scRNA-seq in CRISPR screen. Genome Biol 2020;21:136.
[11] McFarland JM, et al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. Nat Commun 2020;11:4296.
[12] Srivatsan SR, et al. Massively multiplex chemical transcriptomics at single-cell resolution. Science 2020;367:45–51.
[13] Schraivogel D, et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. Nat Methods 2020;17:629–35.
[14] Roth TL, et al. Pooled knockin targeting for genome engineering of cellular immunotherapies. Cell 2020;181:728–744.e21.
[15] Mimitou EP, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods 2019;16:409–12.
[16] Frangieh CJ, et al. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. Nat Genet 2021;53:332–41.
[17] Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. e5 Mol Cell 2017;66:285–99. e5.
[18] Yang L, et al. High-resolution characterization of gene function using single-cell CRISPR tiling screen. Nat Commun 2021;12:4063.
[19] Rubin AJ, et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. Cell 2019;176:361–376.e17.
[20] Pierce SE, Granja JM, Greenleaf WJ. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Nat Commun 2021;12:2969.
[21] Liscovitch-Brauer N, et al. Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. Nat Biotechnol 2021;39:1270–7.
[22] Ursu O, et al. Massively parallel phenotyping of coding variants in cancer with Perturb-seq. Nat Biotechnol 2022;40:896–905.
[23] Otto JE, et al. Structural and functional properties of mSWI/SNF chromatin remodeling complexes revealed through single-cell perturbation screens. Mol Cell 2023;83:1350–1367.e7.
[24] Guna A, et al. A dual sgRNA library design to probe genetic modifiers using genome-wide CRISPRi screens. BMC Genom 2023;24:651.
[25] Sunshine S, et al. Systematic functional interrogation of SARS-CoV-2 host factors using Perturb-seq. Nat Commun 2023;14:6245.
[26] Peidli S, et al. scPerturb: harmonized single-cell perturbation data. Nat Methods 2024. https://doi.org/10.1038/s41592-023-02144-y.
[27] Duan B, et al. Model-based understanding of single-cell CRISPR screening. Nat Commun 2019;10:2233.
[28] Yang L, et al. scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biol 2020;21:19.
[29] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 2019;20:296.
[30] Barry T, Wang X, Morris JA, Roeder K, Katsevich E. SCEPTRE improves calibration and sensitivity in single-cell CRISPR screen analysis. Genome Biol 2021;22:344.
[31] Skinnider MA, et al. Cell type prioritization in single-cell data. Nat Biotechnol 2021;39:30–4.
[32] Nicol, P.B. *et al. Robust Identification of Perturbed Cell Types in Single-Cell RNA-Seq Data* . ⟨http://biorxiv.org/lookup/doi/10.1101/2023.05.06.539326⟩ (2023) ⟨doi:10.1101/2023.05.06.539326⟩.
[33] He B, et al. ASGARD is a single-cell guided pipeline to aid repurposing of drugs. Nat Commun 2023;14:993.
[34] 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;40:245–53.

[35] Papalexi E, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. Nat Genet 2021;53:322–31.
[36] Burkhardt DB, et al. Quantifying the effect of experimental perturbations at single-cell resolution. Nat Biotechnol 2021;39:619–29.
[37] Jin K, et al. CellDrift: inferring perturbation responses in temporally sampled single-cell data. Brief Bioinform 2022;23:bbac324.
[38] Hawkins DY, et al. ICAT: a novel algorithm to robustly identify cell states following perturbations in single-cell transcriptomes. Bioinformatics 2023;39:btad278.
[39] Cao K, Gong Q, Hong Y, Wan L. A unified computational framework for single-cell data integration with optimal transport. Nat Commun 2022;13:7419.
[40] Dong M, et al. Causal identification of single-cell experimental perturbation effects with CINEMA-OT. Nat Methods 2023;20:1769–79.
[41] Suphavilai C, et al. Predicting heterogeneity in clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. Genome Med 2021;13:189.
[42] Peidli S, et al. scPerturb: harmonized single-cell perturbation data. Nat Methods 2024;21:531–40.
[43] Dong X, et al. NetBID2 provides comprehensive hidden driver analysis. Nat Commun 2023;14:2581.
[44] Kamimoto K, et al. Dissecting cell identity via network inference and in silico gene perturbation. Nature 2023;614:742–51.
[45] Kamimoto K, et al. Gene regulatory network reconfiguration in direct lineage reprogramming. Stem Cell Rep 2023;18:97–112.
[46] Jiang, J. *et al. D-SPIN Constructs Gene Regulatory Network Models from Multiplexed scRNA-Seq Data Revealing Organizing Principles of Cellular Perturbation Response*. ⟨http://biorxiv.org/lookup/doi/10.1101/2023.04.19.537364⟩ (2023) ⟨doi:10.1101/2023.04.19.537364⟩.
[47] Ishikawa M, et al. RENGE infers gene regulatory networks using time-series single-cell RNA-seq data with CRISPR perturbations. Commun Biol 2023;6:1290.
[48] Bravo González-Blas C, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods 2023;20:1355–67.
[49] Aibar S, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083–6.
[50] 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.
[51] Yang Y, et al. Gene knockout inference with variational graph autoencoder learning single-cell gene regulatory networks. Nucleic Acids Res 2023;51:6578–92.
[52] Lotfollahi M, Wolf FA, Theis FJ. scGen predicts single-cell perturbation responses. Nat Methods 2019;16:715–21.
[53] Lotfollahi M, Naghipourfar M, Theis FJ, Wolf FA. Conditional out-of-distribution generation for unpaired data using transfer VAE. Bioinformatics 2020;36:i610–7.
[54] Lotfollahi M, et al. Predicting cellular responses to complex perturbations in high-throughput screens. Mol Syst Biol 2023;19:e11517.
[55] Inecik, K., Uhlmann, A., Lotfollahi, M. & Theis, F. MultiCPA: Multimodal Compositional Perturbation Autoencoder. http://biorxiv.org/lookup/doi/10.1101/2022.07.08.499049 (2022) doi:10.1101/2022.07.08.499049.
[56] Seninge L, Anastopoulos I, Ding H, Stuart J. VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. Nat Commun 2021;12:5684.
[57] Kana O, et al. Generative modeling of single-cell gene expression for dose-dependent chemical perturbations. Patterns 2023;4:100817.
[58] Lopez, R. *et al.* Learning Causal Representations of Single Cells via Sparse Mechanism Shift Modeling. Preprint at ⟨http://arxiv.org/abs/2211.03553⟩ (2023).
[59] Wu Y, et al. Predicting cellular responses with variational causal inference and refined relational information. *ICLR* 2023;2023.
[60] Lotfollahi M, et al. Biologically informed deep learning to query gene programs in single-cell atlases. Nat Cell Biol 2023. https://doi.org/10.1038/s41556-022-01072-x.
[61] Bunne, C., Stark, S.G. & Gut, G. Learning Single-Cell Perturbation Responses using Neural Optimal Transport.
[62] Yeo GHT, Saksena SD, Gifford DK. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. Nat Commun 2021;12:3222.
[63] Roohani Y, Huang K, Leskovec J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. Nat Biotechnol 2023. https://doi.org/10.1038/s41587-023-01905-6.
[64] Yu, H. & Welch, J.D. *PerturbNet Predicts Single-Cell Responses to Unseen Chemical and Genetic Perturbations*. ⟨http://biorxiv.org/lookup/doi/10.1101/2022.07.20.500854⟩ (2022) ⟨doi:10.1101/2022.07.20.500854⟩.
[65] Doncevic D, Herrmann C. Biologically informed variational autoencoders allow predictive modeling of genetic and drug-induced perturbations. Bioinformatics 2023;39:btad387.
[66] Zinati, Y., Takiddeen, A. & Emad, A. *GRouNdGAN: GRN-Guided Simulation of Single-Cell RNA-Seq Data Using Causal Generative Adversarial Networks*. ⟨http://biorxiv.org/lookup/doi/10.1101/2023.07.25.550225⟩ (2023) ⟨doi:10.1101/2023.07.25.550225⟩.
[67] Zheng, Y. *et al. Unagi: Deep Generative Model for Deciphering Cellular Dynamics and In-Silico Drug Discovery in Complex Diseases* . ⟨https://www.researchsquare.com/article/rs-3676579/v1⟩ (2023) ⟨doi:10.21203/rs.3.rs-3676579/v1⟩.
[68] Zheng Y, Gao GF. Geneformer: a deep learning model for exploring gene networks. Sci China Life Sci 2023;66:2952–4.
[69] Cui H, 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.
[70] Liu, T., Chen, T., Zheng, W., Luo, X. & Zhao, H. *scELMo: Embeddings from Language Models Are Good Learners for Single-Cell Data Analysis*. ⟨http://biorxiv.org/lookup/doi/10.1101/2023.12.07.569910⟩ (2023) ⟨doi:10.1101/2023.12.07.569910⟩.
[71] Hao, M. *et al. Large Scale Foundation Model on Single-Cell Transcriptomics*. ⟨http://biorxiv.org/lookup/doi/10.1101/2023.05.29.542705⟩ (2023) ⟨doi:10.1101/2023.05.29.542705⟩.
[72] Lopez, R. *et al.* Learning Causal Representations of Single Cells via Sparse Mechanism Shift Modeling. Preprint at ⟨http://arxiv.org/abs/2211.03553⟩ (2023).
[73] Occhipinti A, Verma S, Doan LMT, Angione C. Mechanism-aware and multimodal AI: beyond model-agnostic interpretation. S0962892423002350 Trends Cell Biol 2023. https://doi.org/10.1016/j.tcb.2023.11.002.
[74] Dhainaut M, et al. Spatial CRISPR genomics identifies regulators of the tumor microenvironment. Cell 2022;185:1223–1239.e20.
[75] Rajewsky N, et al. LifeTime and improving European healthcare through cell-based interceptive medicine. Nature 2020;587:377–86.
[76] Ma P, et al. Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. Cell 2023;186:877–891.e14.
[77] Kedzierska, K.Z., Crawford, L., Amini, A.P. & Lu, A.X. Assessing the Limits of Zero-Shot Foundation Models in Single-Cell Biology. http://biorxiv.org/lookup/doi/10.1101/2023.10.16.561085 (2023) doi:10.1101/2023.10.16.561085.
[78] Consens, M.E. *et al.* To Transformers and Beyond: Large Language Models for the Genome. Preprint at ⟨http://arxiv.org/abs/2311.07621⟩ (2023).
[79] Ghandeharioun, A., Caciularu, A., Pearce, A., Dixon, L. & Geva, M. Patchscopes: A Unifying Framework for Inspecting Hidden Representations of Language Models. *ArXiv* **abs/2401.06102**, (2024).
[80] Tejada-Lapuerta, A., Bertin, P., Bauer, S., Aliee, H. & Theis, F.J. Causal machine learning for single-cell genomics.
[81] Palma, A., Theis, F.J. & Lotfollahi, M. *Predicting Cell Morphological Responses to Perturbations Using Generative Modeling*. ⟨http://biorxiv.org/lookup/doi/10.1101/2023.07.17.549216⟩ (2023) ⟨doi:10.1101/2023.07.17.549216⟩.
[82] Caruana R. Multitask learning. Mach Learn 1997;28:41–75.
[83] Tang X, et al. Explainable multi-task learning for multi-modality biological data analysis. Nat Commun 2023;14:2546.
[84] Chandrasekaran SN, et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. Nat Methods 2024. https://doi.org/10.1038/s41592-024-02241-6.
[85] Van de Sande B, et al. Applications of single-cell RNA sequencing in drug discovery and development. Nat Rev Drug Discov 2023;22:496–520.