CellPress

Cell Systems

Perspective

Machine learning for perturbational single-cell omics

Yuge Ji,[1,2] Mohammad Lotfollahi,[1,3] F. Alexander Wolf,[1,4] and Fabian J. Theis[1,2,4,*]
[1]Institute of Computational Biology, Helmholtz Center Munich, Munich, Germany
[2]Department of Mathematics, Technical University of Munich, Munich, Germany
[3]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany
[4]Cellarity, Cambridge, MA, USA
*Correspondence: fabian.theis@helmholtz-muenchen.de
https://doi.org/10.1016/j.cels.2021.05.016

SUMMARY

Cell biology is fundamentally limited in its ability to collect complete data on cellular phenotypes and the wide range of responses to perturbation. Areas such as computer vision and speech recognition have addressed this problem of characterizing unseen or unlabeled conditions with the combined advances of big data, deep learning, and computing resources in the past 5 years. Similarly, recent advances in machine learning approaches enabled by single-cell data start to address prediction tasks in perturbation response modeling. We first define objectives in learning perturbation response in single-cell omics; survey existing approaches, resources, and datasets (https://github.com/theislab/sc-pert); and discuss how a perturbation atlas can enable deep learning models to construct an informative perturbation latent space. We then examine future avenues toward more powerful and explainable modeling using deep neural networks, which enable the integration of disparate information sources and an understanding of heterogeneous, complex, and unseen systems.

INTRODUCTION

Modeling and predicting the effects of perturbations is a key task of systems biology, e.g., for identifying pathway components after knockdowns, cellular response to stimuli, or factors in tissue growth and regeneration. Fundamentally, an accurate understanding of perturbation at the systems level—from binding or docking, to downstream effects and organ-level phenotypes—requires knowledge of the relationships between mechanisms at a molecular, cellular, and tissue level. Perturbation experiments, in which a basal cell-state changes due to an external, controlled event, capture various pieces of these mechanisms, as characterized by biomarkers, cell viability, and drug-protein interactions, to name a few.

Historically, to capture these perturbation mechanisms at scale, various institutes and consortia created systematic databases. *In vitro* drug screens are a common starting point in studying drug effect, so such results and annotations have been collected in databanks such as the Genomics of Drug Sensitivity in Cancer (Yang et al., 2013), the Cancer Therapeutics Response Portal (Basu et al., 2013), and the Connectivity Map (Lamb et al., 2006; Subramanian et al., 2017). Other databases such as PharmacoDB (Smirnov et al., 2016, 2018) and DrugBank (Wishart et al., 2018) do not have data per *in vitro* cell line but per drug, recording general information such as compound targets and chemical properties. To provide a common language for describing compound effects, community efforts striving to "unify biology" curated the gene ontology (GO) (Ashburner et al., 2000), with additional lists of phenotypes available within a landscape of enrichment tools (Huang et al., 2009; Chen et al., 2013; Liberzon et al., 2015; Kuleshov et al., 2016). The combination of these and other databases proved to be useful for many discoveries in perturbation biology through computational and statistical power.

Following this data generation, the past decade of machine learning (ML) for perturbation modeling has used these large-scale data to develop, train, and test modeling approaches. By combining these sources of information, classical ML models have been able to solve tasks such as identification of protein targets and compound-cell line IC_{50} prediction (Rees et al., 2016; Yang et al., 2018b). More recently, deep learning (DL), popular for its flexibility and power to learn complex relationships, has been applied to chemistry and sequence data through convolutions (Gao et al., 2015; Kuenzi et al., 2020), as well as proving fruitful for toxicity prediction from chemical structure (Yang et al., 2018a; Karimi et al., 2019), and cell line sensitivity prediction from somatic mutations (Chang et al., 2018; Chan et al., 2019), and emerged as a resource in perturbation screening (Chandrasekaran et al., 2021). However, elucidating the molecular mechanisms behind perturbation biology have seen less success with deep neural networks, and differences in cell response to perturbations remained poorly understood for the vast majority of compounds.

This manuscript focuses on more recent developments that have emerged as the increasing availability of high-throughput multi-omic data has made it possible to leverage DL methods to establish more fine-grained and predictive models for the above tasks (Zhou and Troyanskaya, 2015; Eraslan et al.,

2019; Zheng and Wang, 2019; Wu et al., 2020). In particular, single-cell profiling methods that generate data with a high number of observations enable training DL models on the transcriptional, proteomic, and epigenetic level, which was previously untenable due to the low number of observations (Hua et al., 2005). This high sample size makes it possible to simultaneously model heterogeneous cellular responses along multiple axes of variation. The foremost example is single-cell RNA sequencing (scRNA-seq), which captures the full heterogeneity of response in hundreds of thousands of cells, augmenting previously low-dimensional data like drug-cell line viability and dose response curves (Goldman et al., 2019; Wu et al., 2020). Because they provide more examples and more variation, single-cell data potentially allow models to extrapolate to unseen events, with fewer experiments. This is specifically made possible by the core DL concept of latent space representation—a distilled representation of the genomic, epigenomic, or proteomic features, which capture the cellular characteristics relevant for understanding and predicting a cell's response to perturbation. This space is constructed as the model identifies and assigns importance to recurring patterns in the data—the co-regulation of a set of factors in a pathway after perturbation, or cell-type markers indicative of cell state, thereby also providing insight into defining characteristics of cells (Way and Greene, 2018). Another way to understand latent space representations is by seeing them as a type of data compression, similar to a PCA, t-SNE (van der Maaten and Hinton, 2008), or UMAP (McInnes et al., 2018) representation but in higher dimensions. For instance, a two-dimensional autoencoder latent space representation is very similar to t-SNE. Generative models (Sohn et al., 2015; Lopez et al., 2018) use this concept to allow implicit modeling of the interaction between cell states and sample (i.e., batch and treatment) effects. Supervised models, which compress the data into a bottleneck, learn patterns of cell-state perturbations.

Despite these advantages, and many approaches to perturbations with microscopic or cellular readouts (Bray et al., 2016; Subramanian et al., 2017; Ye et al., 2018), perturbation modeling and DL in single-cell omics have been largely disparate fields. The most likely contributing factor is a lack of at-scale single-cell perturbation datasets, with a high number of perturbations and/or conditions. DL is data hungry—complex models have difficulty learning generalizable solutions for high-dimensional problems without both many samples and a large number of perturbation conditions (classes). Furthermore, the confounding of batch effect and different experimental variables make it difficult to combine perturbation datasets from separate experiments. However, with appropriate data, perturbation model development can see the same explosion in DL development and performance that followed ImageNet (Russakovsky et al., 2015) in the 2010s, which was driven by both standardized data and benchmarks. With the goal of pushing single-cell perturbation modeling along the same direction, we highlight here similar steps that can be taken. We first propose a general set of objectives of perturbation modeling for single-cell omics in order to contextualize current approaches, as well as datasets that address these objectives. We then discuss future directions of the emerging field, which center around effectively leveraging DL.

PERTURBATION MODELING OBJECTIVES

To provide a framework for summarizing the varied and sometimes scattered landscape of perturbation modeling, we outline four objectives that the field aims to solve (Figure 1). We define an objective as a solvable task, which comes with evaluation metrics that quantify performance. Common themes in examining cellular response to perturbation include categorizing the kinds of cellular states we hope for models to capture, as well as the conditions that constitute a perturbation dataset, both classically and in the more recent context of single-cell data. We address this by asking models to reconstruct and quantify cellular response, and predict drug targets, interactions, and chemical properties (Figure 1). These objectives can serve as guiding categories for scientists embarking on this field. Moreover, they can serve as benchmarking tasks for the ML models themselves and through that, define a common language for establishing validity of methodology contributions to perturbation modeling.

Perturbation response prediction

By predicting the unseen omic signatures and phenotypic measurements in a cell line after perturbation (Figure 1A), a model captures relevant associations of the biological underpinnings of the effect of perturbation. An example of predicting omic signatures (Figure 1Ai) might be predicting RNA expression of a cell type after perturbation. Successful completion can be measured by metrics such as the correlation between real and predicted means and covariances (Lotfollahi et al., 2019, 2020b), with additional importance placed on the prediction of markers. An example of predicting phenotypic measurements (Figure 1Aii) might be predicting IC_{50} values for a cell line and a perturbation, given the IC_{50}s of the perturbation in other cell lines. Successful completion can be measured by classical regression metrics. Such phenotypic measurements are often immediately relevant for drug discovery; for example, drug sensitivity prediction for intratumoral heterogeneity (Shalek and Benson, 2017; Dagogo-Jack and Shaw, 2018; Goldman et al., 2019) may apply when selecting a chemotherapeutic tailored to a patient.

Target and mechanism prediction

Learning the features of and predicting underlying targets and mechanisms (Figure 1B) is the primary way of making perturbation modeling relevant for generating biological understanding (Schenone et al., 2013). An example of this task is predicting the protein targets and pathways activated by a compound. Successful completion can be measured by classification metrics such as precision and recall using the putative pathways and targets for well-characterized compounds. A model that performs well in this task can be used for characterizing novel compounds (Hu and Lill, 2014; Sydow et al., 2019), estimating side effects (Ietswaart et al., 2020; Seo et al., 2020), or repurposing in drug discovery (Seo et al., 2020).

Perturbation interaction prediction

This task covers the broad category of characterizing and predicting the synergy (Meyer et al., 2019) between pairs of perturbations (Figure 1C). Compound or genetic perturbation pairs can either have categorical labels (e.g., synergistic and antagonistic)

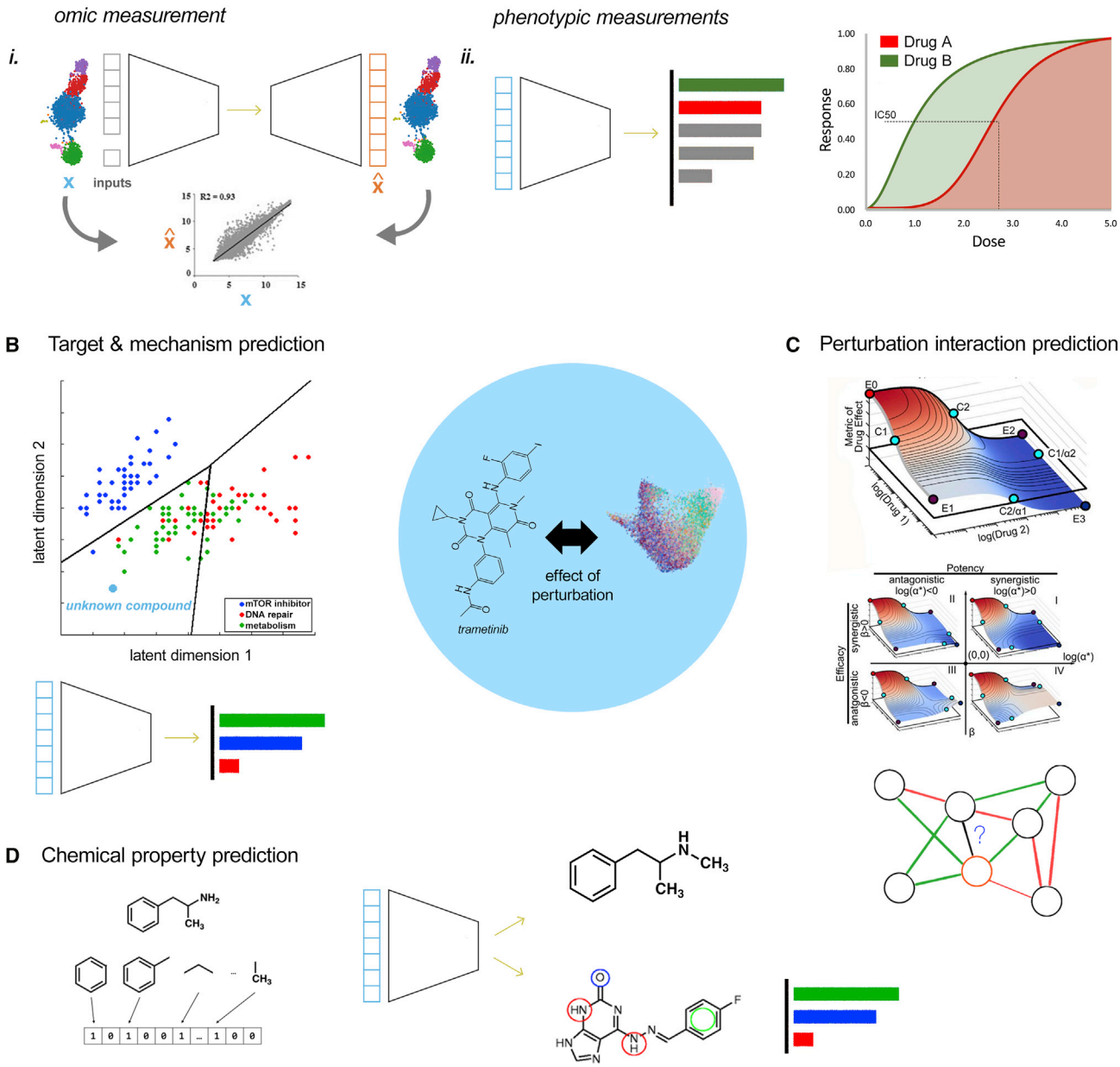

Figure 1. Perturbation modeling objectives as machine learning problems

Here, trapezoid blocks represent any machine learning model, with adjacent light blue rectangles indicating omic data input. Horizontal bar plots indicate instances in which a model output might be categorical, and thus produce a prediction as a ranking from most to least likely.

(A) Perturbation response prediction. (i) *Omics measurements*: predicting the transcriptional or other omic signatures after a perturbation, where the input might contain information about the control condition and variables describing the treatment. The model prediction ($\hat{x}$) can be evaluated using the correlation of features with respect to the true values (X). (ii) *Phenotypic measurements*: predicting continuous scalar descriptors of cell-line response such as IC_{50}, area under the dose response curve, toxicity, or viability.

(B) *Target and mechanism prediction*: predicting canonical targets and mechanisms of perturbations using omic measurements. Each red, blue, and green dot represents a compound in some model representation, which can be used to predict compound MoAs. A model which performs well will be able to assign an MoA even to a previously uncharacterized compound (light blue).

(C) *Perturbation interaction prediction*: predicting combinatorial effects of perturbations, both to understand genetic wiring and synergistic effects. Illustration adapted from Meyer et al., 2019. We can understand this problem as characterizing response curves across pairwise combinations of compounds. The graph depicts an example problem, in which some drugs pairs are synergistic (green), some antagonistic (red), and a model is asked to predict the unknown interaction between two drugs (black).

(D) *Chemical property prediction*: given an omic measurement, predict chemical properties of perturbations, like R groups, molecular fingerprints, pharmacophores, or the complete compound. A model might provide probabilities for which chemical properties the compound contains or predict the compound itself.

or continuous synergy scores, and successful completion can be evaluated with classification or regression metrics, respectively (Menden et al., 2019). This task has been increasingly relevant for developing combination therapies (Bayat Mokhtari et al., 2017). Alternatively, for genetic perturbations, understanding the interactions between genetic targets elucidates network effects. An additional modeling consideration for genetic perturbations is that, unlike in small molecule screens, information from knock up, down, and out experiments (Xiao et al., 2015) can be combined for training of a common model.

Chemical property prediction

Predicting cell-type-unspecific chemical qualities of perturbations based on biological data (Figure 1D) is a primary way to connect biology with chemistry (Camacho et al., 2018). An example is *de novo* generation of small molecules given a transcriptomic profile (Méndez-Lucio et al., 2020), or prediction of classification labels such as R groups and pharmacophores. For the former, successful completion can be measured by the Tanimoto similarity of the *in silico* compound with an existing compound, as well as comparing shared structures. For the latter, standard classification metrics can be used. Connecting biological features to chemistry features is critical for assessing compound effect and toxicity without a need for in vitro experiments. Alternatively, for genetic perturbations, this task is formulated as predicting genetic sequence from omic profiles and might make use of the biological structure, for example, by modeling motifs or k-mers. Such a model would be able to assess genetic perturbation effects without a need for *in vitro* experiments and, as in the third objective (Figure 1C), potentially elucidate network effects.

CURRENT APPROACHES FOR PERTURBATION MODELING IN SINGLE-CELL OMICS

There exists a large body of literature that addresses the above modeling objectives with both systems biology and DL approaches applied on bulk epigenomics, transcriptomics, and proteomics, small molecule phenotypic screens, *in vitro* cell line characterizations, and clinical measurements, which is reviewed, among others, by Patel et al. (2020) and Camacho et al. (2018). In the context of the more recently developed single-cell genomics and mass cytometry methods, the literature is relatively scarce. The latter field (“perturbational scRNA-seq”) first emerged from methods used to analyze data from combined CRISPR and scRNA-seq experiments (Adamson et al., 2016; Dixit et al., 2016; Datlinger et al., 2017). More recently, single-cell perturbation experiments have become more prolific, and perturbation modeling starts to become established as a distinct category in the field, with much expected growth.

We provide a review of existing single-cell omics perturbation models separated into categories based on commonly established ways to categorize ML models (Table 1). All methods except CellOracle (Kamimoto et al., 2020) can be used for perturbations as defined by a before-and-after effect, which could include healthy versus diseased phenotypes (Buschur et al., 2020), or cross-species translation (Lotfollahi et al., 2019; Chen et al., 2020a, 2020b). CellOracle is specific to genetic/single-target perturbations as it infers effect through propagating signal through a gene regulation network (GRN). While MUSIC (Duan et al., 2019) also contains additional tailored functionality for assessing CRISPR single-guide RNA transduction success, the method is not CRISPR screen specific. With the exception of DRUG-NEM, all of the following approaches have available implementations on GitHub, which come with installation instructions and either a tutorial or example usage.

Linear/shallow classification and regression

Linear models learn how combinations of different expression values correspond to various output values, which can be either discrete (e.g., a protein target) or continuous (e.g. IC_{50}). These shallow classifiers and regressors perform well where the amount of data is limited and are common in perturbation modeling outside of single-cell genomics (Bagherian et al., 2021). SCATTome (Mitra et al., 2016) tackles the task of predicting cell-state-specific responses to compounds by using linear regressions, while Augur (Skinnider et al., 2021) takes a similar, albeit nonlinear, approach by using random forest models. Both model how omic profiles can predict a scalar response label per single cell by training (in this case, ensembles of) regression models. MIMOSCA (Dixit et al., 2016) and one of the models in scMAGeCK (Yang et al., 2020) (scMAGeCK-LR) are similarly set up, only that scRNA-seq vectors are the predicted labels and there is no ensembling. MIMOSCA is also not strictly linear as it is a regressor containing interaction terms to model covariates.

Factor decomposition

Factor decomposition methods break down expression values into different components of variation. Matrix factorization has proven popular, both in bulk (Squires et al., 2020; Zhao et al., 2020) and single cell (Mohammadi et al., 2020) genomics analyses, for its interpretability and scalability with larger datasets (Stein-O’Brien et al., 2018); for example, decomposing expression values in scRNA can cluster and rank genes into groups which are readily interpretable as mechanisms, pathways, or processes, most often through GO enrichment. Using this concept of factor decomposition, MUSIC (Duan et al., 2019) groups genes as “topics” through topic modeling and is able to quantify the size of the perturbational effect using the differential activation of all topics. DRUG-NEM (Anchang et al., 2018) calculates a differential probability per proteomic feature per cell and uses the probability matrix into a nested effects model to derive drug combinations.

Nonlinear distribution modeling

Distribution modeling has the advantage of being able to capture the natural variation in biological data. However, these methods are more difficult to interpret. While also applicable to bulk transcriptomics (Umarov and Arner, 2020; Rampášek et al., 2019), distribution modeling gained popularity in the single-cell field as a way to describe population shifts and is especially tractable given the number of cells. PopAlign (Chen et al., 2020a, 2020b) fits a Gaussian and matches perturbed and unperturbed cell populations after factor decomposition into a latent space (with orthogonal non-negative matrix factorization, such that PopAlign is also in part a factor decomposition model). PhEMD

Table 1. Perturbation modeling approaches in single cell omics

				Modeling objectives/prediction types					
Type	Name	Data	Description	a.i	a.ii	b	c	d	Code
Shallow classification and regressions	SCATTome	targeted (qRT-PCR), cell line sensitivity	semi-ensemble of linear predictors	–	x	–	–	–	R, https://github.com/bvnlab/SCATTome
	Augur	scRNA, scATAC, STARmap, MERFISH	measure response as the sampled cross-val AUROC from random forest trained to predict perturbation label	–	x	–	–	–	R, https://github.com/neurorestore/Augur
	MIMOSCA	Perturb-seq	linear regression on combinations with interaction terms for covariates	x	possible	x	x	–	python, https://github.com/asncd/MIMOSCA
Factor modeling	DRUG-NEM	CyTOF	nested effect model using drug-subpopulation effect matrix, maximize combined effect probabilities with fewest compounds	–	–	possible	x	–	not available
	MUSIC	CRISPR screens	topic modeling decomposition of perturbed conditions	–	x	–	–	N/A	R, https://github.com/bm2-lab/MUSIC
	scMAGeCK	CROP-seq	two separate models - LR for perturbations and RRA for ranking genes from the perturbed condition	–	possible	x	–	–	R, https://github.com/weililab/scMAGeCK
Nonlinear distribution modeling	scGen	scRNA	VAE and perturbational effect modeled with latent space arithmetic	x	x	–	–	–	python, https://github.com/theislab/scgen
	trVAE	scRNA	CVAE + maximum mean discrepancy to handle multiple perturbations	x	x	–	–	–	python, https://github.com/theislab/trVAE
	PhEMD	CyTOF, scRNA	earth mover's distance on cluster proportions as embedded by PHATE	x	x	–	x	–	R, https://github.com/KrishnaswamyLab/phemd
	MELD	scRNA	signal processing on the neighbor graph to measure a response per single cell	–	possible	x[b]	–	–	python, https://github.com/KrishnaswamyLab/MELD
	PopAlign	scRNA	alignment of subpopulations represented as Gaussian probability distributions in oNMF latent space	–	x	x[b]	–	–	python, https://github.com/thomsonlab/popalign
	CPA	scRNA, bulkRNA	AE + advarsarial training to learn perturbation and condition embeddings, which can be used for new combinations	x	possible	possible	x	–	python, https://github.com/facebookresearch/CPA
Network models	CellOracle[a]	scRNA, scATAC	signal propagation through inferred gene regulatory networks	x	possible	x	–	N/A	python, https://github.com/morris-lab/CellOracle

The "data" column describes the data modality the model is based on as described in the original paper. We indicate which of the five modeling objectives (Figure 1) the method addresses with a cross (x) or could perform but was not done (possible). scMAGeCK (Yang et al., 2020) is the combination of two separate models and has been categorized as a linear approach as we only discuss scMAGeCK-LR.

[a]For genetic perturbations

[b]Relies on classical differential expression methods

(Chen et al., 2020a, 2020b) fits distributions on clusters and then calculates the differential with Wasserstein distance, while MELD (Burkhardt et al., 2019) fits a distribution on the nearest neighbor graph between cells and assigns each cell a response value. All three of these methods aim to use the difference between distributions to characterize perturbational effect at a higher resolution than would be the case for distributions generated only from unsupervised clusterings. scGen (Lotfollahi et al., 2019) and trVAE (Lotfollahi et al., 2020b) as well as Ghahramani et al. (2018) fit perturbation distributions via DL architectures and represent first ventures into deep learning and generative approaches for predicting treatment effect.

Network models

Network models leverage prior information to construct the relationships between input features. Systems biology has always generated insight by using prior knowledge to contextualize non-exhaustive new information (Yuan et al., 2019). In particular, incorporating prior knowledge can present an opportunity to produce predictive power despite a model seeing zero instances of the prediction objective—for example, predicting drug combinations after only training on samples with individual drugs (Fröhlich et al., 2018). CellOracle is the only model on this list that similarly leverages prior knowledge by constructing gene regulatory networks through an understanding of transcription and transcription factor regulation. CellOracle combines scATAC and scRNA-seq data to create an interpretable network model through which perturbational responses can be predicted, despite the model having seen no perturbational data.

SINGLE-CELL DATASETS FOR PERTURBATION MODELING

The limitations in both performance and achievable tasks of current models owe much to limitations imposed by current data availability. Both the design and performance of ML models depend heavily on the data. Thus, to evaluate the current state of available perturbational data for model development, we compiled seventeen publications, which provide single-cell transcriptomic and proteomic perturbation profiles (Table 2). The majority of datasets are scRNA-seq measurements, and there is no large perturbation dataset generated with single-cell epigenomic profiling.

Despite there being hundreds of single-cell omic datasets (Svensson et al., 2020) and comprehensive expression atlases such as the Human Cell Atlas (Regev et al., 2017), there are relatively few perturbational single-cell screens. Srivatsan et al. (2020) covers by far the highest number of conditions in a single experiment, with 188 conditions across cell lines, doses, and time points (Table 2). McFarland et al. (2020) comes in second with an order of magnitude more cell line characterizations. The vast majority of what serve as perturbation datasets are pooled CRISPR screens with scRNA-seq, which provide an opportunity to obtain transcriptional profiles for hundreds of genetic perturbations at a time and easily allow for combinationatorial perturbations. It is possible that some of these techniques may be combined to more easily generate data with more conditions. In particular, combination experiments outside of genetic perturbations are extremely rare and could potentially reveal much biological insight in combination with previously known negative interactions (Dumbreck et al., 2015) or combination therapies for complex indications (Bayat Mokhtari et al., 2017).

Toward a perturbation atlas

While the advent of sequencing in the past decade has greatly advanced the characterization of biology, it was not until the L1000 assay (Subramanian et al., 2017) that perturbations were characterized transcriptionally at a scale comparable to classical, low-dimensional measurements. Now, the increasing availability of single-cell technologies and in particular, the development of multiplex scRNA-seq via cell hashing (Cao et al., 2017; Stoeckius et al., 2018; McGinnis et al., 2019; Shin et al., 2019; Gehring et al., 2020; Srivatsan et al., 2020) presents an opportunity for a greater breadth of characterization and an understanding of the heterogeneity of perturbation biology (Karen et al., 2020; Dixit et al., 2016). Cellular barcoding can allow hundreds of perturbation treatments to be sequenced in a single experiment (Srivatsan et al., 2020). The full heterogeneous perturbation profiles of complex model systems such as organoids and patient-derived xenografts (Bhimani et al., 2020; Kim et al., 2020) can be captured with single-cell omics that scale to hundreds of thousands of cells (Subramanian et al., 2019; Lukonin et al., 2020; Svensson et al., 2020). As the experimental cost of single-cell measurements decreases, it is increasingly feasible to add more conditions such as more perturbations, time, and dose.

We believe the generation of a ''perturbation atlas''—a characterization of cell types under perturbation that encompasses the broad variety of inducible cell states—will be a significant step forward for basic biology and drug discovery and ascertaining its current limits. It is hypothesized that current atlases will complete cataloging of basal cell states (Camp et al., 2019). We imagine that a perturbation atlas would encompass established catalogs of cell types and tissues (Nieto et al., 2020; Malladi et al., 2015; Regev et al., 2017). Measurements may then be systematically obtained across different conditions. In vitro cell cultures can provide a point of integration with existing databases of annotations. Training ML models on an atlas promises to generate joint latent distributions that describe the possible cellular profiles that can be feasibly characterized (Wagner et al., 2016) (Figure 2) and allows for interpolation and extrapolation to unseen effects. The space of perturbations that is much larger (arguably infinite) can be mapped to a finite space of possible cell states. Finally, we might consider a ''complete'' perturbation atlas to cover cellular phenotypes comprehensively, instead of sampling such that every perturbation and combination of perturbations is tested. Nonetheless, in general, any perturbation atlas will only be complete under the constraint of a fixed set of perturbations.

Key points to keep in mind when characterizing the space of perturbational effect are sources of biological variation and off-target effects. Despite canonical tissue and cell-type categories, biological variation can often result from genetic underpinnings that predispose certain phenotypes, and patient data are a primary source for this variation. Furthermore, frequently in drug discovery, a key point of interest is the off-target effects of perturbations. We view off-target effects as effects of a perturbation but in an unexpected or unintended pathway, cell type, or

Table 2. Perturbation modeling data

Treatment	Method	Source	Omic type	# perturbations	# cell types	# doses	# timepoints	Description	Availability
Genetic targets	Perturb-seq	Dixit et al., 2016	RNA	10, 24	1	–	1–2	TFs followed by LPS treatment in BMDCs, TFs in K562	SCP
	Perturb-seq	Adamson et al., 2016	RNA	9–93 (sgRNA)	1	–	1	contains combinatorial guide delivery	processed
	CRISP-seq	Jaitin et al., 2016	RNA	8–22	1	–	1	TFs, *in vitro* hemato and *in vivo* data	processed
	CROP-seq	Datlinger et al., 2017	RNA	3–29	1–2	-	1	3 experiments, targeting T cell receptors	processed
	CROP-seq	Hill et al., 2018	RNA	32	1	–	1	targeting tumor surpressors in MCF10A with doxorubicin	processed
	CRISPRi	Gasperini et al., 2019	RNA	1,119, 5,779	1	–	1	2 experiments, CRISPRi of enhancer region	processed
	TAP-seq	Schraivogel et al., 2020	RNA	1,778 (enhancers)	1	–	1	targeted enhancers on two chromosomes in K562	processed
	CRISPRa	Norman et al., 2019	RNA	278	1	–	1	induction of gene pair targets, single gene controls	processed
	Perturb-seq	Jost et al., 2020	RNA	25	2	–	1	4 experiments, sgRNA variants with mismatch	processed
	Perturb-seq	Ursu et al., 2020	RNA	200	1	–	1	100 variants each for 2 genes	unavailable
	Perturb-seq	Jin et al., 2020	RNA	35	–	–	1	in vivo mouse brain development	SCP
	perturb-CITE-seq	Frangieh et al., 2021	RNA+protein	248	1	-	1	treatment resistant cancer samples, patient derived	SCP
Small molecules	sci-Plex	Srivatsan et al., 2020	RNA	188	3	4	2	*in vitro* cancer cell lines and small molecules	processed
	multiplexed	Shin et al., 2019	RNA	45	2	1	1	transfected bar codes label perturbation conditions	unavailable
	MIX-seq	McFarland et al., 2020	RNA	1–13	24–99	1	1–5	4 small molecule experiments, one genetic	processed
	CyTOF	Chen et al., 2020a, 2020b	protein	300	1	1	1	breast cancer cells undergoing TGF-β-induced EMT	–
	scRNA-seq	Zhao et al., 2020	RNA	2,6	6,1	–	–	compounds applied to patient resections	processed

Seventeen published single-cell transcriptomic and proteomic datasets with a minimum of 6 perturbations and 30 features. The numbers listed represent a range for the number of different conditions, across the experiments performed for the publication. Control time points are not included in the number of time points listed. The "availability" column describes whether the dataset is publicly available, where SCP is short for Single Cell Portal (https://singlecell.broadinstitute.org/single_cell). Further information such as dataset access points can be found at https://github.com/theislab/sc-pert.

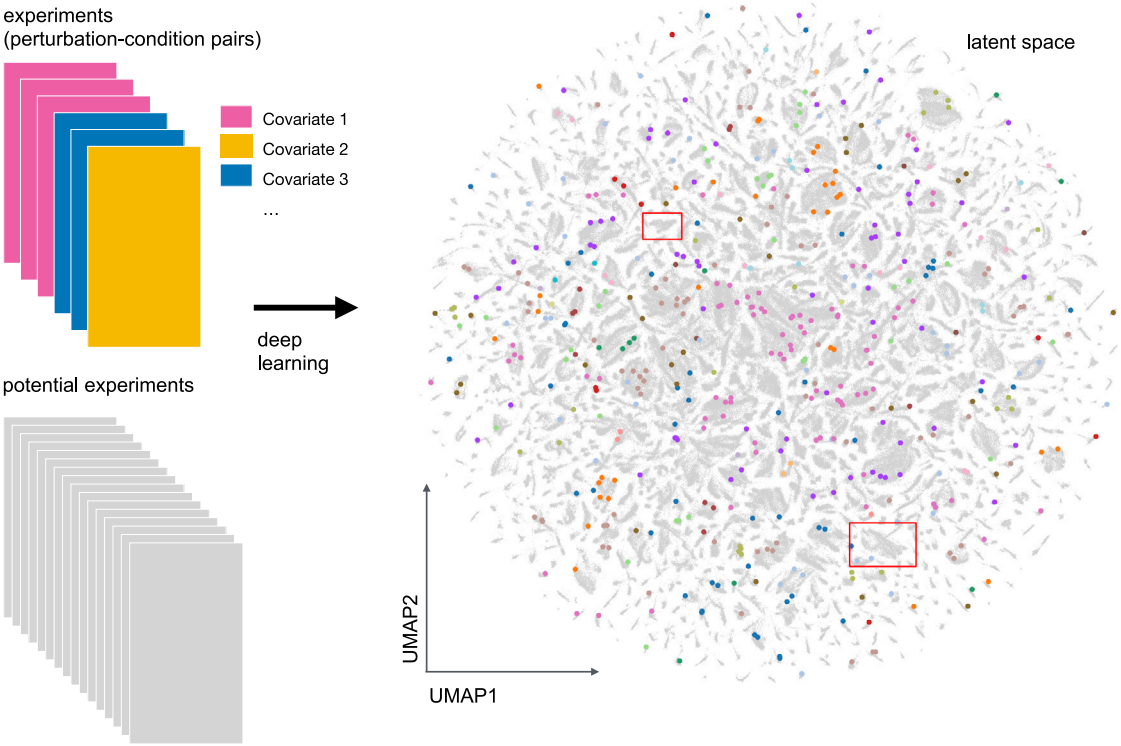

Figure 2. Example schematic of a perturbation latent space in which each dot represents one perturbation-condition pair in the perturbation x condition matrix

Colored dots represent existing experiments and may be related to each other by covariates or labels such as dose, cell type, or pathway activations (Srivatsan et al., 2020). Marked locations represent opportunities for additional sampling.

tissue—as such, a perturbation atlas that provides characterizations for the wide array of compounds and tissues will also provide characterizations for "off-target" effects.

Furthermore, two recent developments make this feasible as a distributed experimental effort: dataset integration (Butler et al., 2018; Haghverdi et al., 2018) mitigates the batch effect downsides of generating the data in separate pieces, and datasets have become easily loadable through single-cell data repositories (Regev et al., 2017; Fischer et al., 2020). An optimized search algorithm may make a large collection even more tenable (Lee et al., 2021). Importantly, training a model with such a compendium of data may require systems infrastructure that is less widely available; this can be circumvented in part by providing smaller, representative data subsets, which resources such as OpenProblems (https://openproblems.bio) seek to accomplish.

To tailor a perturbation atlas to perturbation modeling, it should take into account perturbations and conditions that capture possible cellular states relevant to our objectives: for example, varying response across cell types due to covariates, different targets and mechanisms of perturbations, perturbation interactions (Yofe et al., 2020), and biologically active chemical or sequence features (Becker et al., 2020). Of particular relevance for machine learning and DL, an atlas should capture enough experimental variation so that an ML model can generalize to many unseen situations, in addition to providing a collection of controlled experiments with many shared covariates. This degree of coordination makes it possible to more efficiently and effectively train ML models, interrogate additional conditions, and gain certainty about potential new regulatory mechanisms. To generate such an atlas iteratively, areas with greater uncertainty in the latent representation or lower performance during prediction can be focal points for increased data collection. As confidence increases, it may be possible to annotate and validate using lower-resolution, higher-throughput data.

DL OF PATTERNS IN PERTURBATION BIOLOGY

How can DL come to have a role in understanding biological systems after perturbation? DL achieved breakthroughs in fields such as imaging and genomic sequence data through the availability of large datasets. In the same way, with the many and large single-cell data now being generated (Svensson et al., 2020), which simultaneously profile hundreds of perturbations, it becomes possible for DL models to identify patterns of perturbation effect. Indeed, the flexibility of deep neural networks has often allowed them to outperform classical ML in situations with increasing data size and complexity, also in biology (Wu et al., 2020).

DL methodology is growing rapidly in improved performance, real-world application, and interpretability, which further encourages application to perturbation modeling. As an example of how method development in one field can be transferred to improvements in biology, image classification architectures, which have seen enormous improvements in performance and

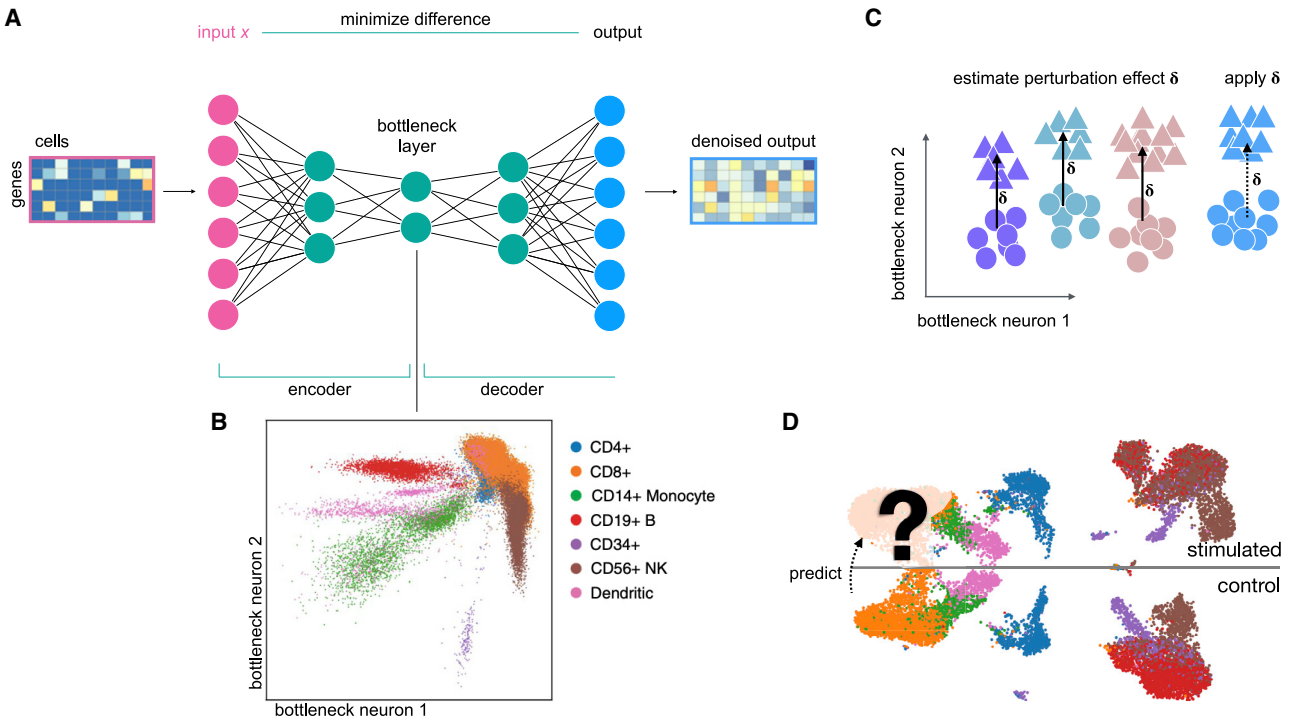

Figure 3. Generative neural network modeling to learn a latent representation of single-cell omics with application to perturbation modeling
(A) An example autoencoder neural network architecture which learns a 2D latent representation, adapted from Eraslan et al. (2019).
(B) Visual representation of the learned latent cell type representation in 2D, illustrating that even in the extreme case of a 2D bottleneck, the network learns a latent representation that preserves cluster structure (unknown to the algorithm), mimicking the common UMAP visualizations.
(C) A schematic of the perturbational effect learned in the latent space representation of scGen, a deep neural network perturbation model that imposes structure via vector arithmetic on the latent space, adapted from Lotfollahi et al., 2019.
(D) Example single perturbation out-of-distribution (i.e., previously unseen) prediction of the response of CD8$^+$ T cells from Lotfollahi et al. (2019), a problem which falls in the category of zero-shot learning.

interpretability (Russakovsky et al., 2015), have been applied to medical images (Ching et al., 2018) and perturbation screens (Cuccarese et al., 2020; Chandrasekaran et al., 2021) with much success in prediction performance. In the following, we outline DL concepts that we expect will contribute to setting up single-cell perturbation models with enough flexibility to address the above key modeling objectives.

Modeling structure with neural networks

Compressing the input data into latent space representations is a key concept in DL for single-cell omics (Lopez et al., 2018). Factorization methods and generative approaches can produce latent structure in an unsupervised fashion. For example, autoencoders—unsupervised neural networks that are trained by funneling data through a bottleneck before reconstructing it—compress data into meaningful components that can be visualized, just as PCA components project data along axes of variation (Figure 3B). We can also use the same models in a supervised manner; latent structure can also be imposed through priors on the condition labels such as vector arithmetic as in Lotfollahi et al. (2019), with maximum mean discrepancy as in Lotfollahi et al. (2020b) (Figure 3C), or more generally with metric learning. When applying these methods, model architectures have been shown to make a large difference: those that take into account existing structure in the data have performed better in learning and compressing data into meaningful latent representations.

Neural network architectures have two well-characterized sub-categories: architectures created to model local feature relationships by using convolution (for example, one pixel is related to the next in image data), or architectures created to model temporal features by creating a model that retains information from previous inputs. These models, which are constructed based on human knowledge and intuition on the task, demonstrate superior performance over fully connected neural networks in capturing complex relationships in image, text, and time series data (Russakovsky et al., 2015; Vaswani et al., 2017; Brown et al., 2020).

Local feature relationships in single-cell omic data

Convolutions are frequently used to detect k-mer patterns in genomic sequence data (Angermueller et al., 2016; Kelley et al., 2018; Movva et al., 2019; Avsec et al., 2021). Similar approaches can be used with single-cell DNA-seq to provide an avenue to investigate cancer treatment response heterogeneity (Luquette et al., 2019; Velazquez-Villarreal et al., 2020), and convolutional DL approaches are candidates for variant identification, predicting minimum residual disease, and tailoring drug options. We envision that the same questions can be asked at the chromosome structure level, in particular given single-cell chromatin conformation capture techniques (Ramani et al., 2017; Lee

et al., 2019). Convolution can also be applied at the intercellular level; spatial transcriptomics make it possible to learn cell-cell neighborhood and interaction patterns, similar to learning pixel-pixel patterns in images. As further example, cell-cell relationships lacking spatial coordinates can be captured in graph convolutions or attention networks (Ravindra et al., 2020). However, while convolutions may be one of the most common methods for feature aggregation, message passing in graph neural networks is a recent advancement, which also performs sparse feature aggregation (Duvenaud et al., 2015) without requiring locality. In all of these methodologies, we will make significant strides in gleaning mechanistic insight and novel biology with perturbation labels (be it small molecules, time, etc.).

Temporal feature relationships in single-cell omic data

In single-cell omic data, amino acid sequences and cells are two inputs that can have a biologically relevant ordering. Inspired by the parallels to language, recurrent neural networks (RNNs) and long short-term memory architectures (LSTMs) have been applied to protein sequences (Bileschi et al., 2019; Karimi et al., 2019). Protein-structure-related methods (Jumper et al., 2020; Du et al., 2017) use neural network architectures heavily influenced by sequence-structure determinants. Another recent DL development, autoencoders architectures combined with attention mechanisms, such as transformers (Vaswani et al., 2017), may contribute to learning recurrent structures in genomic data (Baid, 2018; Clauwaert and Waegeman, 2019; Ji et al., 2020). We believe that what has been learned in these traditionally bulk omics approaches will become relevant for single-cell resolved data, too.

Cells themselves can also be assigned a temporal ordering: recent RNA-labeling methods (Cao et al., 2020; Qiu et al., 2020; Rodriques et al., 2020) would allow us to gather temporal expression data from single-cell perturbation experiments and can be modeled similarly to time series data (Bar-Joseph et al., 2012; Love et al., 2014). Even without RNA labeling, RNA dynamics can also reveal temporal ordering in continuous trajectories (La Manno et al., 2018; Bergen et al., 2020).

Chemical structure priors

No published single-cell perturbation models are yet able to predict chemical properties (modeling objective D), which requires that models incorporate chemical information about perturbations. The lack of such models is attributed in part due to the difficulty of understanding the path to incorporation with omic data (Camacho et al., 2018). However, DL can flexibly combine these sources of information from various single-cell modalities, with chemistry features captured in graph neural networks (Wieder et al., 2020). We anticipate approaches inspired by previous work on encoding chemical structure (Yang et al., 2018a; Méndez-Lucio et al., 2020). Additional priors such as compound docking (Kitchen et al., 2004; Hu and Lill, 2014) could also see use in perturbation characterization, particularly in the case of known targets.

Tabular data

We would like to note at this point that the DL architectures described above are applicable to structured data. However, current single-cell omic data are most often tabular: groups of features (e.g., a set of genes along the same pathway) can determine overall response, but these multicellular programs (Jerby-Arnon and Regev, 2020) differ from context to context, and there is no universal or canonical inherent prior relationships between features (even though graph convolutional networks with properly initialized gene regulatory models may change this in the future). As such, the predominant approach to DL models with single-cell transcriptomic or proteomic input has defaulted to fully connected neural networks, both with large input layers (Dixit et al., 2016; Lopez et al., 2018; Lotfollahi et al., 2019; Brbić et al., 2020; Skinnider et al., 2021) and a high number of features, potentially increasing network complexity in non-beneficial ways. These architectures can be improved by reducing the number of input features with data agnostic approaches such as random projections (Wójcik and Kurdziel, 2019). An advantage of the often shallow, potentially feature-reduced models is easier interpretability than the very deep models, but we expect models to increase in complexity once larger-scale multi-site learning, e.g., learning across the whole Human Cell Atlas (Regev et al., 2017) or Sfaira (Fischer et al., 2020), is more commonly implemented.

Depending on the studied tissue, we believe that a structured approach to single-cell omic data taking advantage of spatial and temporal patterns will be more successful due to the reduced parameters and possible integration of prior knowledge (Lin et al., 2017; Fortelny and Bock, 2020; Gut et al., 2021). While we encourage the application of existing state-of-the-art models for structured data, we believe that intuitions regarding the patterns present in biological, chemical, and perturbation data will lead to cell biology-specific effective DL architectures with fewer and more interpretable input representations.

LEVERAGING NOVEL DL CONCEPTS FOR PERTURBATION LEARNING

The rich toolbox of neural networks offers many building blocks (Goodfellow et al., 2016) that can be transferred to our biological problem, thus enabling integration of novel datasets, other data modalities, reuse of learned models or other tasks of model generalization. Here, we outline these in the context of single-cell perturbation learning.

Multi-task learning

Multi-task learning, in which ML models simultaneously learn to predict multiple labels, can allow models to exploit shared feature relationships between related tasks and learn better regularized latent representations (Camacho et al., 2018), with applications to perturbations (Jiang et al., 2020). Of the methods cited in Table 1, only MIMOSCA is able to address four out of the five model objectives. As the listed objectives draw upon shared interactions between features, we believe multi-task learning will be an important feature of powerful DL models.

Transfer learning

Transfer learning describes taking knowledge learned from one task and applying it to a separate problem, augmenting performance by preserving some previously learned patterns. DL models are particularly amenable for transfer learning due to the ease of carrying over part of the network structure and adapting the rest. Thus, DL has been proven to learn patterns agnostic

of the distribution shifts from dataset to dataset (Gulrajani and Lopez-Paz, 2020; Koh et al., 2020), or to surmount domain shifts through fine-tuning on the target domain (Tajbakhsh et al., 2016; Lotfollahi et al., 2020c). We expect that transferring models (Devlin et al., 2018) learned on a large perturbation atlas (Figure 2) onto a small subset of cellular states together with non-perturbed single-cell atlases will allow models to better generalize to unseen perturbations.

Multi-modal learning

Multi-modal models can simultaneously learn from multiple inputs, often by exploiting the relationships between measurement modalities such as scRNA-seq, ATAC-seq, metabolomics, etc. Current understanding of perturbation biology is extensively grounded in RNA expression, especially as a proxy to protein expression, despite the context dependency of RNA and protein expression correlation (Liu et al., 2016). Furthermore, all of the above methods used a single modality at a time except CellOracle, which was able to generalize to a large number of unseen situations as a result. Thus, while the general integration of multiple modalities in a single model remains an open problem with much potential, we believe modeling should first be guided by biological understanding of how perturbational effect is captured through multiple modalities, and the combination will better inform each modality.

Out-of-distribution prediction

Out-of-distribution learning describes a category of problems in which the model has seen no examples of the relevant class. DL approaches to out-of-distribution or zero-shot learning (Snell et al., 2017; Yu and Lee, 2019) such as generative models (Wang et al., 2017), which act in this latent space regime, have performed especially well (Xian et al., 2017). Because it is impossible to comprehensively screen every perturbation condition combination, this is the predominant problem format of interest in perturbation modeling; for example, a model is trained to characterize a perturbation applied in hundreds of cell types and then asked to predict the effect of perturbation in a new cell type as in scGen. In the latent space of the model, cell types which have similar effects under said perturbation are similar to each other. When a model learns to construct a latent space well, it identifies general features of cell types, which are relevant for predicting its response to perturbation. Leveraging advances in out-of-distribution learning and quantifying when it fails will be key parts of upcoming DL-based perturbation modeling approaches.

Interpretability

A common concern when using DL models is dissecting the reasoning behind model output, especially when building confidence about whether a perturbation model is truly performing well on unseen experiments. However, the interpretability of DL models has seen much progress. Feature attribution methods developed in tasks such as natural language processing or image recognition have proven to generalize to other contexts (Shrikumar et al., 2017; Hooker et al., 2019; Avsec et al., 2020; Lauritsen et al., 2020), and their application to DL perturbation models may be able to elucidate both key transcripts or proteins and their contributions to perturbation response. Disentanglement learning, in which models are constructed to generate latent representations where individual variables represent unique explanatory features, pushes a neural network to dissect the contributions of omics features to perturbation response directly and have been applied to scRNA-seq data (Kimmel, 2020; Lotfollahi et al., 2020a). Additionally, similar to how image data help build intuition through visualizations, interpretability techniques should be validated by constructing ways to discern correctness. Examples of this are DeepLift's *in silico*-generated sequence data with two binding motifs (Shrikumar et al., 2017) and transcription factor enrichment (Stein-O'Brien et al., 2018). We expect similar techniques (applied to e.g., interpret the learned latent space shift in scGen Figure 3C or in other perturbation models) to contribute to understand the convolved latent space structure. Overall, we anticipate that interpretability and outputs with biological relevance are what will distinguish top performers and promote certain methods to become standard practice in analysis pipelines.

CONCLUSION

A perturbation model that is highly predictive captures biologically relevant interactions, even if not obvious at first sight. However, a highly predictive model also serves other ends; successful interpolation and extrapolation to unseen conditions can greatly reduce experimental cost in the case of drug screening, or even make possible patient-specific treatments where data collection is limited. In particular, a perturbation atlas that covers substantial genetic variation might elucidate explanatory features for inter-individual differences in response to perturbation, and genetic differences might be considered a perturbation in and of itself. While the current most cost-effective measure of generating large-scale omic data is via transcriptomics, methods such as single-cell whole-genome sequencing (Gawad et al., 2016) and multi-omic sequencing (Ma et al., 2020; Zachariadis et al., 2020) might provide an opportunity to introduce priors on the relationships between transcriptomic features. Furthermore, perturbation datasets with non-transcriptomic single-cell measurements are scarce, so it is unknown whether different modalities might be better suited for characterizing drug response.

With the expected advent of sufficiently complex perturbation datasets, potentially integrated across cell types and many mapped conditions, we anticipate seeing an explosion of ML methods for modeling single-cell perturbations. Community efforts have already begun, with the DREAM (Bansal et al., 2014; Costello et al., 2014; Menden et al., 2019) and Kaggle challenges producing advances in the classical perturbation modeling framework. The field is further bolstered by several ongoing, potentially highly scalable benchmarking efforts centered separately around single-cell (HCA OpenProblems) and drug discovery (Therapeutic Data Commons; Huang et al., 2021). With more standardized and accessible tasks and data, perturbation modeling in single-cell omics could also become a topic at venues such as the NeurIPS Competition Track. A large-scale public DL challenge with clear objectives and evaluation metrics will push forward imminent breakthroughs in perturbation modeling in the same way AlphaFold2, GPT-3 (Brown et al., 2020), and Inception (Szegedy et al., 2015) set the bar for future methods and encouraged adoption by industry. DL presents an

appealing solution to the multi-label, high-dimensional and multi-modal problems of cellular perturbation characterizations, which are intractable by shallow approaches. We look forward to the integration of network architectures custom-designed to the qualities inherent to biological and chemical data. Now is the time for DL to become a fundamental contributor to single-cell biology.

ACKNOWLEDGMENTS

We thank I. Ibarra for pointing us to protein-target docking algorithms, M. Lueken for pointing us to HCA open problems, and L. Zappia for help with addition of perturbation as a category to scRNA tools and critical comments on the draft of the text. F.J.T. acknowledges support by the the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation, grant #2019- 002438, by the BMBF (grant #01IS18053A and #01IS18036B), and by the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI (grant number ZT-I-PF-5-01) and sparse2big (grant number ZT-I-007).

DECLARATION OF INTERESTS

Y.J. is a consultant for Cellarity and has stake-holder interests. F.A.W. is an employee of Cellarity and has stake-holder interests. F.J.T. reports receiving consulting fees from Cellarity and ownership interest in Cellarity.

REFERENCES

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell *167*, 1867–1882.e21.

Anchang, B., Davis, K.L., Fienberg, H.G., Williamson, B.D., Bendall, S.C., Karacosta, L.G., Tibshirani, R., Nolan, G.P., and Plevritis, S.K. (2018). DRUG-NEM: optimizing drug combinations using single-cell perturbation response to account for intratumoral heterogeneity. Proc. Natl. Acad. Sci. USA *115*, E4294–E4303.

Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O. (2016). Deep learning for computational biology. Mol. Syst. Biol. *12*, 878.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., and Zeitlinger, J. (2020). Base-resolution models of transcription factor binding reveal soft motif syntax. bioRxiv. https://doi.org/10.1101/737981.

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P., and Kelley, D.R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. bioRxiv. https://doi.org/10.1101/2021.04.07.438649.

Bagherian, M., Sabeti, E., Wang, K., Sartor, M.A., Nikolovska-Coleska, Z., and Najarian, K. (2021). Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. Brief. Bioinform. *22*, 247–269. https://doi.org/10.1093/bib/bbz157.

Baid, G. (2018). An attention-based model for transcription factor binding site prediction. https://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-83.pdf.

Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. Nat. Biotechnol. *32*, 1213–1222.

Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. Nat. Rev. Genet. *13*, 552–564.

Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell *154*, 1151–1161.

Bayat Mokhtari, R., Homayouni, T.S., Baluch, N., Morgatskaya, E., Kumar, S., Das, B., and Yeger, H. (2017). Combination therapy in combating cancer. Oncotarget *8*, 38022–38043.

Becker, T., Yang, K., Caicedo, J.C., Wagner, B.K., Dancik, V., Clemons, P., Singh, S., and Carpenter, A.E. (2020). Predicting compound activity from phenotypic profiles and chemical structures. bioRxiv. https://doi.org/10.1101/2020.12.15.422887.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

Bhimani, J., Ball, K., and Stebbing, J. (2020). Patient-derived xenograft models-the future of personalised cancer treatment. Br. J. Cancer *122*, 601–602.

Bileschi, M.L., Belanger, D., Bryant, D., Sanderson, T., Carter, B., Sculley, D., DePristo, M.A., and Colwell, L.J. (2019). Using deep learning to annotate the protein universe. bioRxiv. https://doi.org/10.1101/626507.

Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., and Carpenter, A.E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat. Protoc. *11*, 1757–1774.

Brbić, M., Zitnik, M., Wang, S., Pisco, A.O., Altman, R.B., Darmanis, S., and Leskovec, J. (2020). MARS: discovering novel cell types across heterogeneous single-cell experiments. Nat. Methods *17*, 1200–1206.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., and Askell, A. (2020). Language models are few-shot learners. arXiv. http://arxiv.org/abs/2005.14165.

Burkhardt, D.B., Stanley, J.S., III, Tong, A., Perdigoto, A.L., Gigante, S.A., Herold, K.C., Wolf, G., Giraldez, A.J., van Dijk, D., and Krishnaswamy, S. (2019). Quantifying the effect of experimental perturbations in single-cell RNA-sequencing data using graph signal processing. bioRxiv. https://doi.org/10.1101/532846.

Buschur, K.L., Chikina, M., and Benos, P.V. (2020). Causal network perturbations for instance-specific analysis of single cell and disease samples. Bioinformatics *36*, 2515–2521.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., and Collins, J.J. (2018). Next-generation machine learning for biological networks. Cell *173*, 1581–1592.

Camp, J.G., Platt, R., and Treutlein, B. (2019). Mapping human cell phenotypes to genotypes with single-cell genomics. Science *365*, 1401–1405.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science *357*, 661–667.

Cao, J., Zhou, W., Steemers, F., Trapnell, C., and Shendure, J. (2020). Sci-fate characterizes the dynamics of gene expression in single cells. Nat. Biotechnol. *38*, 980–988.

Chan, H.T., Chin, Y.M., and Low, S.K. (2019). The roles of common variation and somatic mutation in cancer pharmacogenomics. Oncol. Ther. *7*, 1–32.

Chandrasekaran, S.N., Ceulemans, H., Boyd, J.D., and Carpenter, A.E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat. Rev. Drug Discov. *20*, 145–159.

Chang, Y., Park, H., Yang, H.J., Lee, S., Lee, K.Y., Kim, T.S., Jung, J., and Shin, J.M. (2018). Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. Sci. Rep. *8*, 8857.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, S., Rivaud, P., Park, J.H., Tsou, T., Charles, E., Haliburton, J.R., Pichiorri, F., and Thomson, M. (2020a). Dissecting heterogeneous cell populations across drug and disease conditions with PopAlign. Proc. Natl. Acad. Sci. USA *117*, 28784–28794. https://doi.org/10.1073/pnas.2005990117.

Chen, W.S., Zivanovic, N., van Dijk, D., Wolf, G., Bodenmiller, B., and Krishnaswamy, S. (2020b). Uncovering axes of variation among single-cell cancer specimens. Nat. Methods *17*, 302–310.

Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M.M., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface *15*, 20170387. https://doi.org/10.1098/rsif.2017.0387.

Clauwaert, J., and Waegeman, W. (2019). Novel transformer networks for improved sequence labeling in genomics. bioRxiv. https://doi.org/10.1101/836163.

Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. *32*, 1202–1212.

Cuccarese, M.F., Earnshaw, B.A., Heiser, K.A., Fogelson, B., Davis, C.T., McLean, P.F., Gordon, H.B., Skelly, K.R., Weathersby, F.L., Rodic, V., et al. (2020). Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery. bioRxiv. https://doi.org/10.1101/2020.08.02.233064.

Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat. Rev. Clin. Oncol. *15*, 81–94.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. http://arxiv.org/abs/1810.04805.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell *167*, 1853–1866.e17.

Du, X., Sun, S., Hu, C., Yao, Y., Yan, Y., and Zhang, Y. (2017). DeepPPI: boosting prediction of protein-protein interactions with deep neural networks. J. Chem. Inf. Model. *57*, 1499–1510.

Duan, B., Zhou, C., Zhu, C., Yu, Y., Li, G., Zhang, S., Zhang, C., Ye, X., Ma, H., Qu, S., et al. (2019). Model-based understanding of single-cell CRISPR screening. Nat. Commun. *10*, 2233.

Dumbreck, S., Flynn, A., Nairn, M., Wilson, M., Treweek, S., Mercer, S.W., Alderson, P., Thompson, A., Payne, K., and Guthrie, B. (2015). Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines. BMJ *350*, h949.

Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., and Adams, R.P. (2015). Convolutional networks on graphs for learning molecular fingerprints. arXiv. http://arxiv.org/abs/1509.09292.

Eraslan, G., Avsec, Ž., Gagneur, J., and Theis, F.J. (2019). Deep learning: new computational modelling techniques for genomics. Nat. Rev. Genet. *20*, 389–403.

Fischer, D.S., Dony, L., König, M., Moeed, A., Zappia, L., Tritschler, S., Holmberg, O., Aliee, H., and Theiset, F.J. (2020). Sfaira accelerates data and model reuse in single cell genomics. bioRxiv. https://doi.org/10.1101/2020.12.16.419036.

Fortelny, N., and Bock, C. (2020). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. Genome Biol. *21*, 190.

Frangieh, C.J., Melms, J.C., Thakore, P.I., Geiger-Schuller, K.R., Ho, P., Luoma, A.M., Cleary, B., Jerby-Arnon, L., Malu, S., Cuoco, M.S., et al. (2021). Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. Nat. Genet. *53*, 332–341.

Fröhlich, F., Kessler, T., Weindl, D., Shadrin, A., Schmiester, L., Hache, H., Muradyan, A., Schütte, M., Lim, J.-H., Heinig, M., et al. (2018). Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. Cell Syst. *7*, 567–579.e6.

Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y., et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat. Med. *21*, 1318–1325.

Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell *176*, 377–390.e19.

Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. *17*, 175–188.

Gehring, J., Hwee Park, J., Chen, S., Thomson, M., and Pachter, L. (2020). Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. Nat. Biotechnol. *38*, 35–38.

Ghahramani, A., Watt, F.M., and Luscombe, N.M. (2018). Generative adversarial networks simulate gene expression and predict perturbations in single cells. bioRxiv. https://doi.org/10.1101/262501.

Goldman, S.L., MacKay, M., Afshinnekoo, E., Melnick, A.M., Wu, S., and Mason, C.E. (2019). The impact of heterogeneity on single-cell sequencing. Front. Genet. *10*, 8.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning (MIT Press).

Gulrajani, I., and Lopez-Paz, D. (2020). In search of lost domain generalization. arXiv. http://arxiv.org/abs/2007.01434.

Gut, G., Stark, S.G., Rätsch, G., and Davidson, N.R. (2021). PmVAE: Learning interpretable single-cell representations with pathway modules. bioRxiv. https://doi.org/10.1101/2021.01.28.428664.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Hill, A.J., McFaline-Figueroa, J.L., Starita, L.M., Gasperini, M.J., Matreyek, K.A., Packer, J., Jackson, D., Shendure, J., and Trapnell, C. (2018). On the design of CRISPR-based single-cell molecular screens. Nat. Methods *15*, 271–274.

Hooker, S., Erhan, D., Kindermans, P.-J., and Kim, B. (2019). A benchmark for interpretability methods in deep neural networks. In Advances in Neural Information Processing Systems (Curran Associates), pp. 9737–9748.

Hu, B., and Lill, M.A. (2014). PharmDock: a pharmacophore-based docking program. J. Cheminform. *6*, 14.

Hua, J., Xiong, Z., Lowey, J., Suh, E., and Dougherty, E.R. (2005). Optimal number of features as a function of sample size for various classification rules. Bioinformatics *21*, 1509–1515.

Huang, da W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang, K., Fu, T., Gao, W., Zhao, Y., Roohani, Y., Leskovec, J., Coley, C.W., Xiao, C., Sun, J., and Zitnik, M. (2021). Therapeutics data Commons: machine learning datasets and tasks for therapeutics. arXiv. http://arxiv.org/abs/2102.09548.

Ietswaart, R., Arat, S., Chen, A.X., Farahmand, S., Kim, B., DuMouchel, W., Armstrong, D., Fekete, A., Sutherland, J.J., and Urban, L. (2020). Machine learning guided association of adverse drug reactions with in vitro target-based pharmacology. EBiomedicine *57*, 102837.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell. https://doi.org/10.1016/j.cell.2016.11.039.

Jerby-Arnon, L., and Regev, A. (2020). Mapping multicellular programs from single-cell profiles. bioRxiv. https://doi.org/10.1101/2020.08.11.245472.

Ji, Y., Zhou, Z., Liu, H., and Davuluri, R.V. (2020). DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. bioRxiv. https://doi.org/10.1101/2020.09.17.301879.

Jiang, Y., Rensi, S., Wang, S., and Altman, R.B. (2020). DrugOrchestra: jointly predicting drug response, targets, and side effects via deep multi-task learning. bioRxiv. https://doi.org/10.1101/2020.11.17.385757.

Jost, M., Santos, D.A., Saunders, R.A., Horlbeck, M.A., Hawkins, J.S., Scaria, S.M., Norman, T.M., Hussmann, J.A., Liem, C.R., Gross, C.A., et al. (2020). Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. Nat. Biotechnol. *38*, 355–364.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Tunyasuvunakool, K., Ronneberger, O., Bates, R., Žídek, A., Bridgland, A., et al. (2020). High accuracy protein structure prediction using deep learning. In Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book). https://predictioncenter.org/casp14/doc/CASP14_Abstracts.pdf.

Kamimoto, K., Hoffmann, C.M., and Morris, S.A. (2020). CellOracle: dissecting cell identity via network inference and in silico gene perturbation. bioRxiv. https://doi.org/10.1101/2020.02.17.947416.

Karen, V., Assefa, A.T., Yigit, N., Anckaert, J., Vandamme, N., Rombaut, D., Saeys, Y., Thas, O., Speleman, F., Durinck, K., and Vandesompele, J. (2020). Comprehensive benchmarking of single cell RNA sequencing technologies for characterizing cellular perturbation. bioRxiv. https://doi.org/10.1101/2020.11.25.396523.

Karimi, M., Wu, D., Wang, Z., and Shen, Y. (2019). DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. Bioinformatics *35*, 3329–3338.

Kelley, D.R., Reshef, Y.A., Bileschi, M., Belanger, D., McLean, C.Y., and Snoek, J. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res. *28*, 739–750.

Kim, J., Koo, B.K., and Knoblich, J.A. (2020). Human organoids: model systems for human biology and medicine. Nat. Rev. Mol. Cell Biol. *21*, 571–584.

Kimmel, J.C. (2020). Disentangling latent representations of single cell RNA-seq experiments. bioRxiv. https://doi.org/10.1101/2020.03.04.972166.

Kitchen, D.B., Decornez, H., Furr, J.R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov. *3*, 935–949.

Koh, P.W., Sagawa, S., Marklund, H., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., et al. (2020). WILDS: a Benchmark of in-the- wild distribution shifts (arXiv). http://arxiv.org/abs/2012.07421.

Kuenzi, B.M., Park, J., Fong, S.H., Sanchez, K.S., Lee, J., Kreisberg, J.F., Ma, J., and Ideker, T. (2020). Predicting drug response and synergy using a deep learning model of human cancer cells. Cancer Cell *38*, 672–684.e6. https://doi.org/10.1016/j.ccell.2020.09.014.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929–1935.

Lauritsen, S.M., Kristensen, M., Olsen, M.V., Larsen, M.S., Lauritsen, K.M., Jørgensen, M.J., Lange, J., and Thiesson, B. (2020). Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat. Commun. *11*, 3852.

Lee, D.S., Luo, C., Zhou, J., Chandran, S., Rivkin, A., Bartlett, A., Nery, J.R., Fitzpatrick, C., O'Connor, C., Dixon, J.R., and Ecker, J.R. (2019). Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat. Methods *16*, 999–1006.

Lee, J.T.H., Patikas, N., Kiselev, V.Y., and Hemberg, M. (2021). Fast searches of large collections of single-cell data using scfind. Nat. Methods *18*, 262–271.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Lin, C., Jain, S., Kim, H., and Bar-Joseph, Z. (2017). Using neural networks for reducing the dimensions of single-cell RNA-Seq data. Nucleic Acids Res. *45*, e156.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. Cell *165*, 535–550.

Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058.

Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. Nat. Methods *16*, 715–721.

Lotfollahi, M., Dony, L., Agarwala, H., and Theis, F. (2020a). Out-of-distribution prediction with disentangled representations for single-cell RNA sequencing data. Proceedings of the paper workshop on computational biology, ICML 2020 Workshop on Computational Biology (WCB). https://www.researchgate.net/publication/344243171_Out-of-distribution_prediction_with_disentangled_representations_for_single-cell_RNA_sequencing_data.

Lotfollahi, M., Naghipourfar, M., Theis, F.J., and Wolf, F.J. (2020b). Conditional out-of-distribution generation for unpaired data using transfer VAE. Bioinformatics *36*, i610–i617.

Lotfollahi, M., Naghipourfar, M., Malte, D., Luecken, M.D., Khajavi, M., Büttner, M., Avsec, Z., Misharin, A.V., and Theis, F.J. (2020c). Query to reference single-cell integration with transfer learning. bioRxiv. https://www.biorxiv.org/content/10.1101/2020.07.16.205997v1.abstract.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lukonin, I., Serra, D., Challet Meylan, L., Volkmann, K., Baaten, J., Zhao, R., Meeusen, S., Colman, K., Maurer, F., Stadler, M.B., et al. (2020). Phenotypic landscape of intestinal organoid regeneration. Nature *586*, 275–280.

Luquette, L.J., Bohrson, C.L., Sherman, M.A., and Park, P.J. (2019). Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. Nat. Commun. *10*, 3908.

Ma, A., McDermaid, A., Xu, J., Chang, Y., and Ma, Q. (2020). Integrative methods and practical challenges for single-cell multi-omics. Trends Biotechnol. *38*, 1007–1022.

Malladi, V.S., Erickson, D.T., Podduturi, N.R., Rowe, L.D., Chan, E.T., Davidson, J.M., Hitz, B.C., Ho, M., Lee, B.T., Miyasato, S., et al. (2015). Ontology application and use at the ENCODE DCC. Database (Oxford) *2015*, bav010. https://doi.org/10.1093/database/bav010.

McFarland, J.M., Paolella, B.R., Warren, A., Geiger-Schuller, K., Shibue, T., Rothberg, M., Kuksenko, O., Colgan, W.N., Jones, A., Chambers, E., et al. (2020). Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. Nat. Commun. *11*, 4296.

McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat. Methods *16*, 619–626.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. arXiv. http://arxiv.org/abs/1802.03426.

Menden, M.P., Wang, D., Mason, M.J., Szalai, B., Bulusu, K.C., Guan, Y., Yu, T., Kang, J., Jeon, M., Wolfinger, R., et al. (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat. Commun. *10*, 2674.

Méndez-Lucio, O., Baillif, B., Clevert, D.A., Rouquié, D., and Wichard, J. (2020). De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. Nat. Commun. *11*, 10.

Meyer, C.T., Wooten, D.J., Paudel, B.B., Bauer, J., Hardeman, K.N., Westover, D., Lovly, C.M., Harris, L.A., Tyson, D.R., and Quaranta, V. (2019). Quantifying drug combination synergy along potency and efficacy axes. Cell Syst. *8*, 97–108.e16.

Mitra, A.K., Mukherjee, U.K., Harding, T., Jang, J.S., Stessman, H., Li, Y., Abyzov, A., Jen, J., Kumar, S., Rajkumar, V., and Ness, B.V. (2016). Single-cell analysis of targeted transcriptome predicts drug sensitivity of single cells within human myeloma tumors. Leukemia *30*, 1094–1102.

Mohammadi, S., Davila-Velderrain, J., and Kellis, M. (2020). A multiresolution framework to characterize single-cell state landscapes. Nat. Commun. *11*, 5399.

Movva, R., Greenside, P., Marinov, G.K., Nair, S., Shrikumar, A., and Kundaje, A. (2019). Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays. PLoS One *14*, e0218073.

Nieto, P., Elosua-Bayes, M., Trincado, J.L., Marchese, D., Massoni-Badosa, R., Salvany, M., Henriques, A., Mereu, E., Moutinho, C., and Ruiz, S. (2020). A single-cell tumor immune atlas for precision oncology. bioRxiv. https://doi.org/10.1101/2020.10.26.354829.

Norman, T.M., Horlbeck, M.A., Replogle, J.M., Ge, A.Y., Xu, A., Jost, M., Gilbert, L.A., and Weissman, J.S. (2019). Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science *365*, 786–793.

Patel, L., Shukla, T., Huang, X., Ussery, D.W., and Wang, S. (2020). Machine learning methods in drug discovery. Molecules *25*. https://doi.org/10.3390/molecules25225277.

Qiu, Q., Hu, P., Qiu, X., Govek, K.W., Cámara, P.G., and Wu, H. (2020). Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. Nat. Methods *17*, 991–1001.

Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z., and Shendure, J. (2017). Massively multiplex single-cell Hi-C. Nat. Methods *14*, 263–266.

Rampášek, L., Hidru, D., Smirnov, P., Haibe-Kains, B., and Goldenberg, A. (2019). Dr.Vae: improving drug response prediction via modeling of drug perturbation effects. Bioinformatics *35*, 3743–3751.

Ravindra, N., Sehanobish, A., Pappalardo, J.L., Hafler, D.A., and van Dijk, D. (2020). Disease state prediction from single-cell data using graph attention networks. Proceedings of the ACM conference on health, inference, and learning, 121–130.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The human cell atlas. eLife *6*. https://doi.org/10.7554/eLife.27041.

Rodriques, S.G., Chen, L.M., Liu, S., Zhong, E.D., Scherrer, J.R., Boyden, E.S., and Chen, F.((2020). RNA timestamps identify the age of single molecules in RNA sequencing. Nat. Biotechnol. *39*, 320–325.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. *115*, 211–252.

Schenone, M., Dančík, V., Wagner, B.K., and Clemons, P.A. (2013). Target identification and mechanism of action in chemical biology and drug discovery. Nat. Chem. Biol. *9*, 232–240.

Schraivogel, D., Gschwind, A.R., Milbank, J.H., Leonce, D.R., Jakob, P., Mathur, L., Korbel, J.O., Merten, C.A., Velten, L., and Steinmetz, L.M. (2020). Targeted Perturb-seq enables genome-scale genetic screens in single cells. Nat. Methods *17*, 629–635.

Seo, S., Lee, T., Kim, M.H., and Yoon, Y. (2020). Prediction of side effects using comprehensive similarity measures. BioMed Res. Int. *2020*, 1357630.

Shalek, A.K., and Benson, M. (2017). Single-cell analyses to tailor treatments. Sci. Transl. Med. *9*. https://doi.org/10.1126/scitranslmed.aan4730.

Shin, D., Lee, W., Lee, J.H., and Bang, D. (2019). Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. Sci. Adv. *5*, eaav2249.

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning important features Through propagating activation differences. arXiv http://arxiv.org/abs/1704.02685.

Skinnider, M.A., Squair, J.W., Kathe, C., Anderson, M.A., Gautier, M., Matson, K.J.E., Milano, M., Hutson, T.H., Barraud, Q., Phillips, A.A., et al. (2021). Cell type prioritization in single-cell data. Nat. Biotechnol. *39*, 30–34. https://doi.org/10.1038/s41587-020-0605-1.

Smirnov, P., Kofia, V., Maru, A., Freeman, M., Ho, C., El-Hachem, N., Adam, G.A., Ba-Alawi, W., Safikhani, Z., and Haibe-Kains, B. (2018). PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. Nucleic Acids Res. *46*, D994–D1002.

Smirnov, P., Safikhani, Z., El-Hachem, N., Wang, D., She, A., Olsen, C., Freeman, M., Selby, H., Gendoo, D.M., Grossmann, P., et al. (2016). PharmacoGx: an R package for analysis of large pharmacogenomic datasets. Bioinformatics *32*, 1244–1246.

Snell, J., Swersky, K., and Zemel, R.S. (2017). Prototypical networks for few-shot learning. arXiv. http://arxiv.org/abs/1703.05175.

Sohn, K., Lee, H., and Yan, X. (2015). Learning structured output representation using deep conditional generative models. Adv. Neural Inf. Process. Syst. *28*, 3483–3491.

Squires, C., Shen, D., Agarwal, A., Shah, D., and Uhler, C. (2020). Causal imputation via synthetic interventions. arXiv. http://arxiv.org/abs/2011.03127.

Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J., Packer, J., Pliner, H.A., Jackson, D.L., Daza, R.M., Christiansen, L., et al. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. Science *367*, 45–51.

Stein-O'Brien, G.L., Arora, R., Culhane, A.C., Favorov, A.V., Garmire, L.X., Greene, C.S., Goff, L.A., Li, Y., Ngom, A., Ochs, M.F., et al. (2018). Massively multiplex chemical transcriptomics at single-cell resolution. Trends Genet. *34*, 790–805.

Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell *171*, 1437–1452.e17.

Subramanian, A., Sidhom, E.H., Emani, M., Vernon, K., Sahakian, N., Zhou, Y., Kost-Alimova, M., Slyper, M., Waldman, J., Dionne, D., et al. (2019). Single cell census of human kidney organoids shows reproducibility and diminished off-target cells after transplantation. Nat. Commun. *10*, 5462.

Svensson, V., da Veiga Beltrame, E., and Pachter, L. (2020). A curated database reveals trends in single-cell transcriptomics. Database (Oxford) *2020*. https://doi.org/10.1093/database/baaa073.

Sydow, D., Burggraaff, L., Szengel, A., van Vlijmen, H.W.T., IJzerman, A.P., van Westen, G.J.P., and Volkamer, A. (2019). Advances and challenges in computational target prediction. J. Chem. Inf. Model. *59*, 1728–1742.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision. arXiv. http://arxiv.org/abs/1512.00567.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., and Liang., J. (2016). Convolutional neural networks for medical image analysis: full training or fine tuning? arXiv. http://arxiv.org/abs/1706.00712.

Umarov, R., and Arner, E. (2020). A DeepFake framework for prediction of cell type specific transcriptional states induced by drug treatment. bioRxiv. https://doi.org/10.1101/2020.12.14.422792.

Ursu, O., Neal, J.T., Shea, E., Thakore, P.I., Jerby-Arnon, L., Nguyen, L., Dionne, D., Diaz, C., Bauman, J., Mosaad, M., et al. (2020). Massively parallel phenotyping of variant impact in cancer with Perturb-seq reveals a shift in the spectrum of cell states induced by somatic mutations.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. arXiv. http://arxiv.org/abs/1706.03762.

Velazquez-Villarreal, E.I., Maheshwari, S., Sorenson, J., Fiddes, I.T., Kumar, V., Yin, Y., Webb, M.G., Catalanotti, C., Grigorova, M., Edwards, P.A., et al. (2020). Single-cell sequencing of genomic DNA resolves sub-clonal heterogeneity in a melanoma cell line. Commun. Biol. *3*, 318.

Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. Nat. Biotechnol. *34*, 1145–1160.

Wang, W., Pu, Y., Verma, V.K., Fan, K., Zhang, Y., Chen, C., Rai, P., and Carin, L. (2017). Zero-shot learning via class-conditioned deep generative models. arXiv. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewFile/16087/16709.

Way, G.P., and Greene, C.S. (2018). Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. Pac. Symp. Biocomput. *23*, 80–91.

Wieder, O., Kohlbacher, S., Kuenemann, M., Garon, A., Ducrot, P., Seidel, T., and Langer, T. (2020). A compact review of molecular property prediction with graph neural networks. Drug Discov. Today Technol. https://doi.org/10.1016/j.ddtec.2020.11.009.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. *46*, D1074–D1082.

Wójcik, P.I., and Kurdziel, M. (2019). Training neural networks on high-dimensional data using random projection. Pattern Anal. Applic. *22*, 1221–1231.

Wu, Z., Lawrence, P.J., Ma, A., Zhu, J., Xu, D., and Ma, Q. (2020). Single-cell techniques and deep learning in predicting drug response. Trends Pharmacol. Sci. *41*, 1050–1065. https://doi.org/10.1016/j.tips.2020.10.004.

Xian, Y., Lampert, C.H., Schiele, B., and Akata, Z. (2017). Zero-shot learning – a comprehensive evaluation of the good, the bad and the ugly. arXiv. https://doi.org/10.1145/219717.219748.

Xiao, Y., Gong, Y., Lv, Y., Lan, Y., Hu, J., Li, F., Xu, J., Bai, J., Deng, Y., Liu, L., et al. (2015). Gene Perturbation Atlas (GPA): a single-gene perturbation repository for characterizing functional mechanisms of coding and non-coding genes. Sci. Rep. *5*, 10889.

Yang, H., Sun, L., Li, W., Liu, G., and Tang, Y. (2018a). In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. Front. Chem. *6*, 30.

Yang, L., Zhu, Y., Yu, H., Cheng, X., Chen, S., Chu, Y., Huang, H., Zhang, J., and Li, W. (2020). scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biol. *21*, 19.

Yang, M., Simm, J., Lam, C.C., Zakeri, P., van Westen, G.J.P., Moreau, Y., and Saez-Rodriguez, J. (2018b). Linking drug target and pathway activation for effective therapy using multi-task learning. Sci. Rep. *8*, 8322.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

Ye, C., Ho, D.J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R., Henault, M., Mostacci, N., Farmer, P., Renner, S., et al. (2018). DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nat. Commun. *9*, 4307.

Yofe, I., Dahan, R., and Amit, I. (2020). Single-cell genomic approaches for developing the next generation of immunotherapies. Nat. Med. *26*, 171–177.

Yu, H., and Lee, B. (2019). Zero-shot learning via simultaneous generating and learning. arXiv. https://papers.nips.cc/paper/2019/file/19ca14e7ea6328a42e0eb13d585e4c22-Paper.pdf.

Yuan, B., Shen, C., Luna, A., Korkut, A., Marks, D.S., Ingraham, J., and Sander, C. (2019). Interpretable machine learning for perturbation biology. bioRxiv. https://doi.org/10.1101/746842.

Zachariadis, V., Cheng, H., Andrews, N., and Enge, M. (2020). A highly scalable method for joint whole-genome sequencing and gene-expression profiling of single cells. Mol. Cell *80*, 541–553.e5.

Zhao, W., Dovas, A., Spinazzi, E.F., Levitin, H.M., Upadhyayula, P., Sudhakar, T., Marie, T., Otten, M.L., Sisti, M., Bruce, J.N., Canoll, P., and Sims, P.A. (2020). Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. bioRxiv. https://doi.org/10.1101/2020.04.22.056341.

Zheng, J., and Wang, K. (2019). Emerging deep learning methods for single-cell RNA-seq data analysis. Quant. Biol. *7*, 247–254.

Zhou, J., and Troyanskaya, O.G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. Nat. Methods *12*, 931–934.