PORTLAND PRESS

Review Article

An expanded role for single-cell chemical genomics profiling in drug discovery

Adeya Wyatt[1], Kevin Hoffer-Hawlik[1,2], Ross M. Giglio[3], Elham Azizi[1,2,4,5,6] and José L. McFaline-Figueroa[1,2,5]

[1]Department of Biomedical Engineering, Columbia University, New York, NY 10027, U.S.A.; [2]Irving Institute for Cancer Dynamics, Columbia University, New York, NY 10027, U.S.A.; [3]Department of Molecular Pharmacology and Therapeutics, Columbia University Medical Center, New York, NY 10032, U.S.A.; [4]Department of Computer Science, Columbia University, New York, NY 10027, U.S.A.; [5]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, U.S.A.; [6]Data Science Institute, Columbia University, New York, NY 10027, U.S.A.

Correspondence: José L. McFaline-Figueroa (jm5200@columbia.edu)

OPEN ACCESS

The integration of single-cell genomics into the chemical genetics paradigm is reshaping how researchers profile drug activity, prioritize lead candidates, and uncover new therapeutic opportunities. Traditional chemical genetic approaches, though instrumental in linking compounds to cellular phenotypes, often rely on bulk measurements that obscure important cellular heterogeneity and limit insight into mechanisms of action. By contrast, single-cell technologies offer a transformative view of how compounds influence diverse cell types and states, capturing nuanced molecular responses that further our understanding of efficacy, resistance, and polypharmacology. From cancer to neurodegenerative disorders and other disease contexts, single-cell chemical profiling enables a more precise annotation of drug-induced effects, revealing differential responses across cellular subpopulations. These methods help identify both beneficial and adverse outcomes that may not be readily predicted by a compound's structure or known targets, enhancing preclinical prioritization and supporting rational drug repurposing strategies. As these technologies mature, advances in multiplexing, multimodal profiling, and computational analysis are expanding their scalability and applicability to increasingly complex models. The resulting data-rich assays are poised to bridge critical gaps between compound screening and clinical relevance. This review highlights the evolution of chemical genomics toward single-cell resolution and outlines emerging opportunities to leverage these methods throughout the drug discovery pipeline, from early preclinical prioritization to late-stage repurposing, ultimately accelerating the development of safer, more effective therapies.

Introduction

The advent of genomic technologies has significantly enhanced our ability to catalog how perturbations influence cellular states [1–6]. In the context of disease, genomics tools can help identify actionable disease drivers [7–9], map genetic variation to cellular phenotypes [10–13], and define regulatory networks associated with therapeutic efficacy [14–16]. More recent advances applying these technologies at single-cell and spatial resolution have further refined our understanding of disease, pinpointing specific cell types and states underlying disease phenotypes [17] and highlighting a basis for heterogeneity in disease manifestation [18].

Chemical genetics revolutionized our ability to dissect how thousands of compounds and biologics alter gene function and cellular phenotypes [19]. High-throughput chemical genetic screens applied across genetically diverse cellular models, both naturally occurring [2] and engineered through genetic manipulation [20], have helped define the genetic dependencies of cellular responses to various exposures. Many genetic screens are limited to gross phenotypic (such as viability and proliferation) or highly specific molecular readouts (such as transcriptional reporters and enzymatic activity). While valuable, these approaches often limit insight into mechanisms of action or off-target effects. Over the past two decades, the integration of large-scale genomic profiling into chemical genetics has offered a more comprehensive understanding of cellular responses to perturbations and laid the foundation for the extension of this field into chemical genomics [21].

Drug discovery strategies are broadly categorized into target-based drug discovery (TDD) and phenotypic drug discovery (PDD) [22]. In TDD, a molecular target is nominated a priori based on biological or genetic evidence, and screening efforts focus on identifying compounds or biologics optimized for their ability to modulate that target. The TDD approach is analogous to reverse genetic screens in which perturbation of a defined gene is used to interrogate downstream phenotypes [19]. In PDD, compounds are prioritized based on their ability to elicit a cellular phenotype of interest, without

requiring prior knowledge of the underlying target. This approach parallels forward genetic screens, which begin from a phenotype and work backward to identify causal genes. In the context of chemical genomics, target-based campaigns use high-throughput, information-rich profiling and computational tools to understand how modulation of a target reshapes biological states [23]. In phenotypic screens, it enables the deconvolution of targets and pathways underlying active compounds by linking exposure to detailed molecular response profiles [24].

In drug discovery, whether in target-based or phenotypic campaigns, there is an inherent balance between employing well-established, low-information-content assays that are specific, rapid, and scalable [25] and leveraging unbiased, information-rich profiling [24]. However, drug discovery efforts may lack a comprehensive annotation of the diverse molecular effects induced by a compound. This gap can result in unanticipated drug-induced effects, an incomplete understanding of a drug's mechanism of action, and poor clinical translation [26–28]. By systematically integrating large-scale profiling of drug-induced molecular effects alongside standard targeted assays that define on-target activity, researchers can improve the characterization of a drug's mechanism of action and its often complex (poly-)pharmacology.

Within the drug development pipeline, chemical genomics tools present a unique opportunity to enhance the characterization of compound libraries, promising lead candidates, and both investigational and approved drugs based on their ability to induce distinct molecular states. Applying this paradigm early in the pipeline would enable the identification of beneficial and detrimental effects that may not be apparent from a compound's chemical structure, class, or annotated mechanism of action. Such insights have the potential to improve prioritization, mitigate unforeseen negative effects, and refine therapeutic strategies.

In this review, we explore the methodological and computational advances that can be readily applied to the drug discovery pipeline. Specifically, we highlight areas where high-throughput single-cell chemical genomics profiling can drive progress in compound identification, preclinical prioritization, and the characterization of a drug's polypharmacology, as well as facilitate drug repurposing efforts (Figure 1). For a more detailed discussion on target identification, disease-specific drug discovery, genomics applications, quality control, and best practices, we refer readers to the following reviews [29–33].

Population-averaged chemical genomics approaches

Before the rapid advancement of next-generation sequencing (NGS), high-throughput parallel measures of gene expression first took place on DNA microarrays [34], capturing relative mRNA levels using hybridization to gene-specific probes. Marton et al. first established the idea of a chemical 'signature,' or a gene expression footprint, demonstrating not only on-target calcineurin signaling inhibition-induced expression signatures in *Saccharomyces cerevisiae* but also off-target *GCN4*-regulated transcriptional responses [3]. Later, Hughes et al. built on this with the idea of collecting a 'compendium' of signatures, showing in *S. cerevisiae* that deletions in genes involved in similar pathways result in more similar signatures compared with gene deletions involved in separate pathways [35]. This idea proved its value as the authors assigned function to several unannotated yeast ergosterol pathway genes; their search narrowed substantially simply by inspecting how a signature for a gene of unknown function clustered with those of known genes, and the authors suggest the future use of the approach for chemical perturbations [35].

The compendium approach was applied in what would become the Connectivity Map (CMap), part of the National Institutes of Health (NIH) Common Fund's Library of Integrated Network-based Cellular Signatures (LINCS) program [36]. Originally a collection of 453 perturbation-induced expression signatures across four human cancer cell lines measured by DNA microarrays [37], the subsequent CMap studies measure perturbation response by L1000, an assay using bead hybridization and fluorescent measurements of 978 landmark transcripts as a proxy for the expression of the transcriptome [2]. The approach has allowed for the profiling of a large number of perturbations with the latest iteration containing 1.16M replicate-collapsed signatures from >80 k perturbations (33,609 compounds; 9288 genes) across >200 cell models [38]. By connecting signatures of uncharacterized chemicals to those of known chemicals and genetic perturbation via gene loss-of-function, CMap authors classified previously unknown ROCK and MTOR/PI3K inhibitors and discovered the first inhibitor against CSNK1A1. Resources at this scale have the potential to enable discovery of effective therapeutics, especially when coupled with orthogonal resources like the Cancer Dependency Map [1,39,40], an encyclopedic endeavor to uncover genetic dependencies across an ever-growing collection of pan-cancer cell lines, including the extensive genomic characterization of these cell lines [41–43], and examining viability changes in response to a
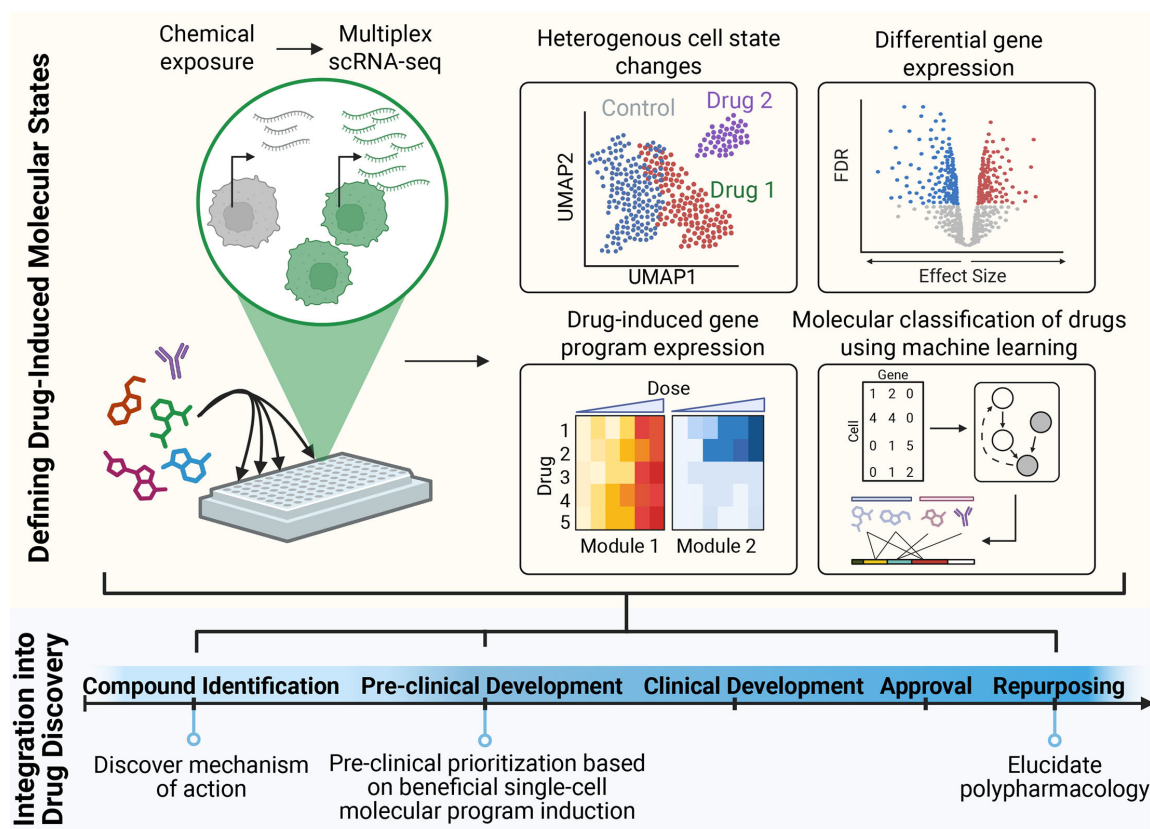
Figure 1: High-throughput single-cell chemical genomics.

Results can drive progress in compound identification, preclinical prioritization, and the characterization of a drug's polypharmacology, as well as facilitate drug repurposing efforts.

diverse set of drugs [44,45]. Notably, the L1000 assay, while advantageous for its low cost and scale, is limited by the number of genes measured, requiring the imputation of 9196 of 11,350 unmeasured genes. The authors note that missing information from the L1000 readout could contribute to the inability to connect 37% of the screened compounds to their expected target and that higher sensitivity measurements should be considered for future studies.

More recently, with the lowering costs of sequencing and transcriptome profiling, bulk RNA-seq chemical genomic approaches were developed to address the challenge of faithfully measuring the whole transcriptome while maintaining high-throughput capacity. Landmark studies, PLATE-seq [4] and later DRUG-seq [46], established multi-well plate-based bulk RNA-seq protocols to measure genome-wide expression changes in response to chemical or genetic perturbation by leveraging well-specific reverse transcription barcodes. The PLATE-seq and DRUG-seq studies went on to collect signatures from 184 and 433 chemicals, respectively, demonstrating the feasibility of collecting unbiased transcriptomic responses to study the molecular response to chemicals from a mechanistic perspective.

In addition to gene expression profiling, other groups have enabled assays with orthogonal genomic measurements of drug response, such as epigenomics and proteomics. With regard to the latter, efforts to measure protein-based drug signatures [47] have resulted in the discovery and further characterization of HDAC inhibitors [48,49], JNK inhibitors [50], adrenergic receptor antagonists [51], CDK inhibitors [52], and BRAF inhibitors [53], to name several examples, providing mechanistic insights into drug–protein interaction, intra-protein stability and disruption, and interactions within protein complexes [54].

Population-averaged, or bulk, approaches highlight the utility of collecting and comparing perturbation-induced gene expression signatures. From these data and with follow-up validation, researchers were able to assign function to uncharacterized genes, classify or reclassify chemicals by mechanism of action, and identify off-target inhibition. These previous studies laid the foundation for the application of large-scale chemical genomics and continue to be resources in forming hypotheses for

drug discovery and development. Over the last two decades, advances in single-cell genomics technologies have led to the discovery of striking heterogeneity in cell type composition and an astounding diversity in cellular state across normal and pathological conditions [55–60]. Population-averaged approaches, with their limited sensitivity to detect these differences, average over heterogeneity, often leading to a muted view of induced changes and, in some instances, can reverse the true trends in the data, a phenomenon known as Simpson's paradox [61]. Of relevance to chemical genomics, bulk averaging can mask the response of clinically relevant cellular subpopulations and ignore heterogeneity in drug-induced molecular responses. Lastly, despite atlas-scale efforts, it is nearly impossible for a single resource to characterize perturbation signatures of interest to researchers (i.e. provide data for the inhibitor of interest in a particular model). Therefore, accessible screening methods, technologies that account for heterogeneity, and computational tools capable of inferring drug-induced effects are a necessity to uncover unique genotype- and therapy-specific transcriptional responses.

Chemical genomics at single-cell resolution

Population-averaged, bulk chemical genomics approaches coupled with high-throughput screens led to a significant advance in the ability to perform molecular characterization within the drug-discovery pipeline. However, these methods are still limited by (1) the inability to capture heterogeneous changes in cell state, which can frequently occur even within seemingly homogeneous cellular systems, and (2) the lack of resolution in complex heterogeneous systems compatible with high-throughput screening (e.g. organoids) to monitor cell type and cell state specific responses and how changes in cell type composition correlate with drug response.

In the past 15 years, single-cell omics techniques have revolutionized our ability to identify the cellular basis of disease pathology and aid in the prioritization of therapeutic approaches. Leveraging the resolution of single-cell omics technologies is quickly enabling the discovery of translationally impactful correlations between precise cell-state measurements and disease, which have the opportunity to feed back to advance drug discovery. Within the purview of single-cell chemical genomics, the systematic molecular screening of drugs at single-cell resolution has the potential to expedite the drug discovery pipeline. Similar to bulk methods, it can enable annotations that inform treatments with increased efficacy and minimize detrimental effects, uncovering mechanisms to combat drug resistance all while determining whether these properties vary across cellular subpopulations. For example, in the context of cancer, single-cell transcriptomics approaches have been applied to gain insight into the resistance continuum [62], the gradual adaptation of cancer cells to cancer therapy with increasing fitness. Single-cell profiling of a *BRCA2*-deficient high-grade serous ovarian cancer cell line with varying concentrations of treatment identified vulnerabilities to PARP inhibition with olaparib through the combinatorially targeting of glutaminase 1 with CB-839, increasing treatment efficacy compared with monotherapy PARP inhibition treatment, and inducing synthetic lethality in *BRCA1*- and *BRCA2*-deficient tumors [62]. Single-cell genomics technologies have similarly been applied to define how canonical amyloid-β and tau pathologies regulate microglia subtypes during Alzheimer disease progression through scRNA-seq [63] and sex-specific gene-expression changes in neurodegenerative diseases that can be monitored by profiling cell types of the blood [64].

The large number of conditions within a high-throughput chemical screen brings several challenges to the use of single-cell molecular readouts, in particular, the need for technologies with sufficient throughput and the need to minimize technical variation across batches. Thresholds for efficiency and cost-effectiveness should also be considered when choosing an appropriate readout. Currently, there are various single-cell genomics approaches that have or can be readily adapted to chemical genomic profiling (nanowell, droplet-based microfluidic, and combinatorial indexing/split-pool techniques), each with differing characteristics in sensitivity, cost, and throughput. We first discuss the general characteristics of methods that can be applied to routinely capture at least thousands of cells per experiment, focusing on the measure of cellular transcriptomes and how they are suitable for single-cell chemical genomics profiling, followed by introducing a variety of multiplexing strategies at the exposure or cellular level that can minimize technical noise when applied to large screens and can be leveraged to increase the throughput of these techniques. Lastly, we comment on how further innovation on these platforms has enabled multi-modal profiling approaches at the level of the proteome, epigenome, and genome (Figure 2).

Nano-well-based strategies (e.g. Seq-Well [65] and ICELL8 [66]) use gravity (sedimentation) to physically isolate individual cells into nanowell arrays, resulting in an efficient, simple, low-cost system.
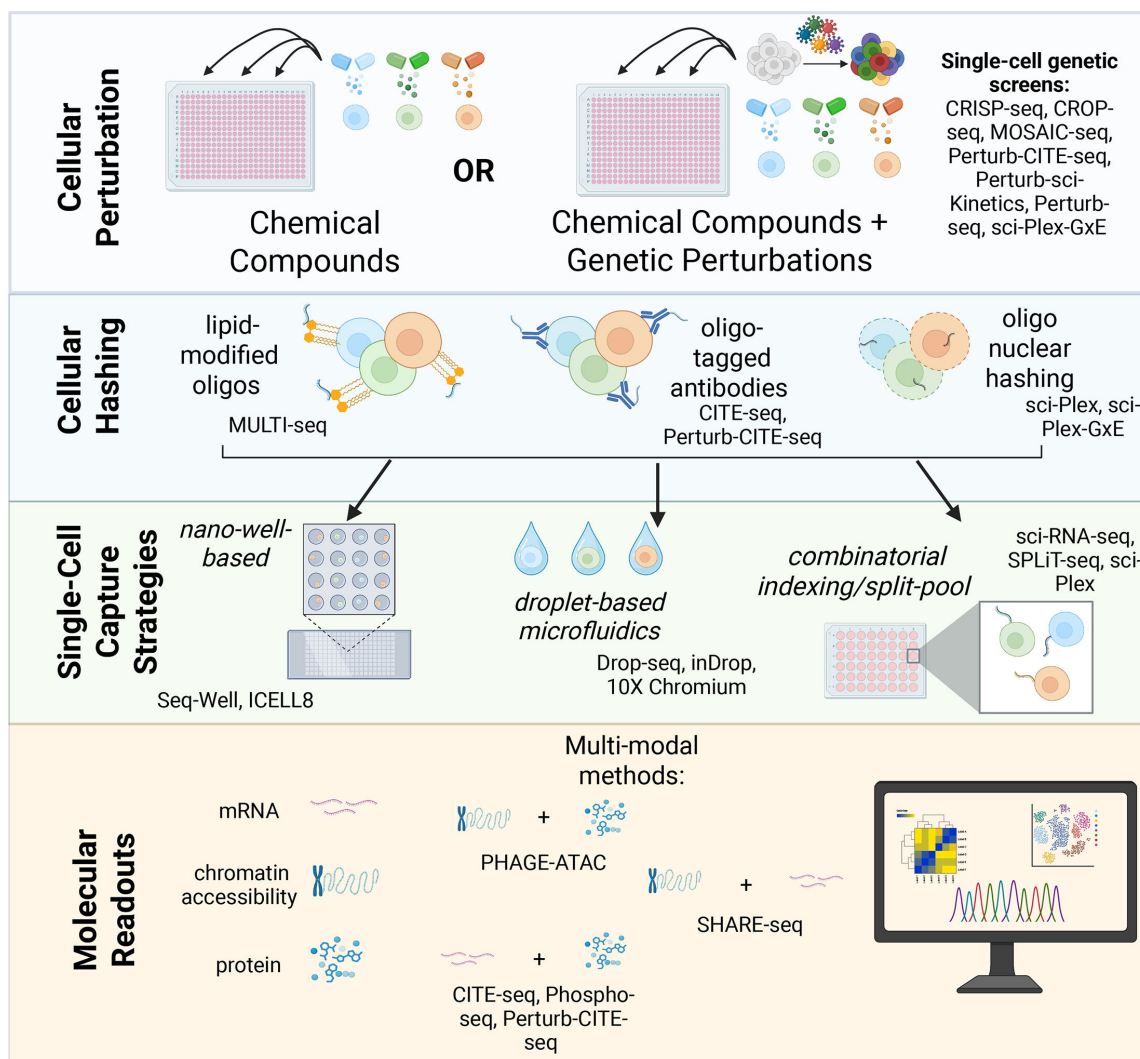
Figure 2: An overview of the general workflow of wet lab methods for chemical genomics screening.

After exposing cells to chemical compounds, various hashing cellular and nuclear hashing methods may be used before applying one of the three single-cell genomics strategies: nano-well based, droplet-based microfluidics, or combinatorial indexing/split-pool. Finally, the high-throughput, multiplexed data produced by these techniques are computationally analyzed, with some methods enabling multi-modal insights.

Despite having a lower overall throughput than the droplet-based microfluidic and combinatorial indexing/split-pool techniques described below, nanowell-based methods have a high capture rate for a cell's transcriptome and are ideal for low cell number samples [67]. The scalability and native throughput of nanowell-based methods, however, are limited by the number of array wells, although cellular hashing techniques, described later, could allow for array 'superloading' to increase throughput in single-cell chemical genomics screens.

Droplet-based microfluidic techniques are the most commonly used single-cell methods for high-throughput single-cell scRNA-seq. In this approach, individual cells and barcoded beads are co-encapsulated within an aqueous/oil emulsion within nanoliter droplets that uniquely tag a cellular transcriptome. Drop-seq [68] and inDrop [69] constitute the first iterations of these methods. Drop-seq, developed by Macosko and colleagues, separates single cells into droplets, capturing single-cell transcriptomes attached to microparticles (STAMPs) for reverse transcription, amplification, and sequencing using the STAMP barcodes to identify the cell of origin. InDrop (indexing droplets) encapsulates cells with a reverse transcription mix and hydrogel beads containing primers released by UV irradiation. Prior to sequencing, this method employs T7 RNA polymerase for linear amplification via *in vitro* transcription. The commercialization of droplet-microfluidic approaches (e.g. the Chromium

system from 10X Genomics) led to an exponential increase in the accessibility of researchers to single-cell RNA-seq technology and the innovation of novel assays on the platform. The throughput of commercial droplet-based microfluidic techniques allows for the capture of up to 10–20,000 cells per unique sample. Cellular hashing approaches, described below, can be used to increase this throughput by overloading nanoliter droplets. However, these approaches have limits on the size of cells that can be captured (large cells may block the device) and can include technical noise and background associated with ambient RNA [65].

Combinatorial indexing or split/pool methods (e.g. sci-RNA-seq [70] and SPLiT-seq [71]) constitute highly scalable single-cell genomics technologies. In contrast to other methods, cells are not physically isolated for library preparation, but rather single cells or nuclei are uniquely labeled via a combination of barcodes across several indexing rounds. Throughput scales rapidly with the number of indexing rounds allowing for the profiling of millions of cells within one experiment. The approaches have allowed for comprehensive profiling of millions of cells representing cell types and cell states across multiple organisms [72,73]. Similar to microfluidic approaches, a key consideration is that the number of available cellular barcodes is significantly larger than the number of cells, minimizing barcode collisions [74]. In the context of drug discovery, this level of scale pairs well with the notion of large throughput screens and was demonstrated in our development of the sci-Plex pipeline for multiplex single-cell chemical transcriptomics [5]. Recently, the suitability of these technologies for large-scale single-cell chemical genomics profiling was excitingly highlighted by the release of the Tahoe-100M, a 100 million cell dataset, collected using a split-pool technique, of mosaic collections of cancer cell lines exposed to hundreds of compounds at various drug/dose combinations [75].

Drug-induced cellular states can be characterized by changes in RNA transcription, chromatin accessibility, protein expression, metabolisms, or other molecular changes. Any one technology (genomics, transcriptomics, proteomics, and metabolomics) fails to capture the full scope of a single cell's identity, making multi-modal multiplexed methods ideal for deeper insights into cellular drug response. Multiple technologies have been developed to measure additional aspects of the central dogma at single-cell resolution, frequently in combination with other measurements [[76–82]]. For example, CITE-seq and Phospho-seq profile both the transcriptome and quantify surface or intracellular proteins with oligonucleotide-labeled antibodies [83,84]. Assays to profile chromatin accessibility profiling, such as the assay for transposase accessible chromatin (ATAC-seq [85]) have been extended to droplet-based microfluidics [86] and combinatorial indexing [87] approaches for gene expression analysis. PHAGE-ATAC measures phage-based multiplex protein measurements and chromatin accessibility through scATAC-seq in parallel, utilizing phages for epitope profiling and phage libraries to select antigen-specific libraries [88]. SHARE-seq and sci-CAR simultaneously measure chromatin accessibility and gene expression, linking accessibility to transcription across cell states [89,90]. DEFND-seq co-profiles mRNA and genomic DNA (gDNA) from an individual cell's nucleus, building on ATAC-seq by revealing nucleosome positioning [91]. NEAT-seq [92] reduces non-specific staining of barcoded antibodies to nuclear proteins by conjugating *Escherichia coli* ssDNA binding protein, enabling simultaneous profiling of intra-nuclear proteins, chromatin accessibility, and gene expression. The epigenome has also been studied at single-cell resolution using DNA methylation profiling via bisulfite conversion with high-throughput single-cell combinatorial indexed techniques using sciMETv2 [93], where non-methylated cytosine bases are converted into uracil and sequenced as thymine. Lastly, techniques for whole genome sequencing at single-cell resolution (sc-WGS) and the direct capture of primary DNA sequence to profile mutations and areas of genetic variation (e.g. SMOOTH-seq [94]) provide the final piece of the central dogma. Exciting next steps for the fields include the incorporation of metabolic measurements to arrive at comprehensive measures of cellular state. Single-cell metabolic regulome profiling (scMEP) [95], for example, an antibody-based, single-cell metabolic profiling method, has been used to define the metabolic states of human cytotoxic T cells. While techniques that simultaneously assess the genome, transcriptome, and proteome of an individual cell together have not been widely established, several techniques do provide multi-modal insight that leads to more detailed cell profiling, better potential drug-target identification, and broader drug response prediction. Systems such as Element Bioscience's AVITI24 are working to close this gap by combining NGS, RNA capture, proteomics, and morphology of a sample in one platform [96].

Cellular hashing approaches for single-cell genomics allow multiple samples, such as cellular models exposed to diverse chemical libraries, to be pooled and processed jointly within one experiment, increasing throughput, minimizing batch effects, and reducing costs [97]. Most approaches involve the addition of unique DNA-barcoded tags assigned to each sample and synthetic analytes to be captured within

each methodology. These tags enable pooled samples to be deconvolved in downstream analysis and individual cells to be matched to specific conditions. A range of cellular hashing methods has been optimized and implemented. One of the first examples of cellular hashing was developed in 2018 by Stoeckius et al. as a modification of the multiplexed single-cell multi-modal measurement technique CITE-seq [98]. This multiplexing strategy allows for the increase of parallel experiments and helps identify cell multiplets signaling by using a defined set of oligo-tagged antibodies to attach to proteins on the cell surface. MULTI-seq [99], a technique that specifically tags cells with lipids and cholesterol-conjugated oligonucleotides to barcode live cells and nuclei. Using MULTI-seq, McGinnis et al. identified transcriptional responses of myoepithelial and luminal epithelial cells across co-culture conditions and after induction of transforming growth factor beta (TGF-β) signaling. The combinatorial indexing-based sci-Plex approach, taking advantage of the inherent affinity of DNA molecules to the cell nucleus, involves attaching single-stranded DNA oligonucleotide barcodes to permeabilized cells. This low-cost multiplexing strategy coupled to combinatorial indexing sci-RNA-seq3 was used in an unbiased single-cell chemical genomics screen profiling thousands of unique combinations of cell model, drug, and dose, identifying a novel metabolic basis of the molecular response of cells to histone deacetylase inhibition [5] and classifying related classes of small molecule inhibitors (EGFR inhibitors) by their ability to induce distinct molecular programs [100].

The tracking of genomic heterogeneity within captured cellular transcripts also allows for multiplexing across diverse cellular models. Demuxlet [101] and Mix-seq [102] use sample-specific demultiplexing based on genetic single-nucleotide polymorphism to deconvolve mixtures of cell types. Mix-seq demonstrates the ability to use this approach to multiplex conditions across diverse cell models exposed to differing drugs and multiple time points [102]. Lastly, the incorporation of heritable barcodes amenable to capture via single-cell genomics techniques can be used for multiplexing and to track the evolution of cells under therapy. CellTag Indexing [103] tracks cells and cellular lineages across different conditions over different time points by the expression of randomized molecular indexes within RNA polymerase II driven transcripts via sequential lentiviral delivery and has recently been extended to multimodal profiling [104]. Cheng and colleagues recently used CellTag to generate a lentivirus barcode library and observed an association between the induction of resistance to targeting of the kinesin motor KIF11 by ispinesib in glioblastoma (GBM) cells and a transition to the improved-survival-associated proneural glioma subtype [105].

The addition of single-cell genetic screens within the context of the chemical genomic pipelines allows for the ability to systematically dissect which genes are causative for drug-induced molecular changes. Beginning in 2016, the development of the Perturb-seq [106,107], CRISP-seq [108], CROP-seq [109], and MOSAIC-seq [110] approaches, which couple CRISPR-based genetic perturbations with a single-cell transcriptomics readout, revolutionized our ability to determine how gene activity alters gene expression networks. The first iteration of these methods relied on capture of CRISPR single guide RNA (sgRNA) identity by capturing a polyadenylated transcript containing a distal barcode that reports on the delivered sgRNA or embedding of the sgRNA within the 3′ untranslated region (UTR) of a reporter transcript in droplet-microfluidics based single-cell RNA-seq. Distal barcode methods can suffer from decreased statistical power due to lentiviral recombination during pooled lentivirus generation [111], which can be minimized by creating viral particles individually and pooling for transduction [107]. More recently, innovations have allowed for direct capture of short sgRNAs [112,113], which fully mitigates challenges associated with distal barcodes. Recently, the sci-Plex-GxE [114] and Perturb-sci-Kinetics [115] frameworks have demonstrated the ability to couple single-cell genetic screens with high-throughput combinatorial indexing RNA-seq, applying them to identify the regulation of resistance programs downstream of drug-induced transcription and the regulation of RNA synthesis, respectively.

Advancements in the scalability of high-throughput screens, multiplexing strategies, and cellular hashing have led to the generation of large, complex datasets. Despite the fact that data are becoming increasingly multiplexed as the scope of single-cell analysis increases, the single-cell chemical genomics field is limited by both technique efficiency and older demultiplexing analysis. To truly make use of single-cell chemical genomics and gain biological insight to be applied to drug discovery, innovation is still needed to produce more robust readouts while still facilitating scale. Newer computational techniques are also needed in order to extract meaningful information from these sparse, large-scale screens regarding cellular response to drug exposure. To elucidate more mechanisms of action, better dissect immune response, and unbiasedly catalog drugs with single-cell chemical genomics, more advancements must be made in computational frameworks to analyze increasingly complex, high-throughput single-cell data.

Computational methods to gain insight from highly multiplex chemical genomics screens

The applications of large-scale chemical genomics screens to drug discovery require the development of novel computational tools that account for considerations specific to multiplex chemical screens and allow us to gain insight from these extraordinarily high-dimensional datasets while accounting for confounding technical artifacts of single-cell experiments. As discussed in [29], the adoption of high-throughput drug screens with molecular readouts has already revolutionized drug discovery, allowing for the identification of therapeutic cellular targets and a better understanding of drug mechanism of action. Further, these assays are well poised to generate the necessary data toward the rational design of drug combinations with predicted therapeutic effects and to inform drug repurposing efforts.

Advances in machine learning (ML) methods applied to scRNA-seq to learn and detect perturbational response patterns have accompanied the generation of single-cell perturbation datasets [116]. A key challenge is that perturbation responses are typically measured on different cells before vs. after treatment, yielding unpaired single-cell data distributions. Another crucial objective in perturbation modeling is out-of-distribution prediction, or predicting treatment effects from unseen conditions, given that it is infeasible to explore the full combinatorial space of drug perturbations. A model must first learn perturbation-induced transcriptional effects and subsequently, the model can predict transcriptional effects from counterfactual conditions, including untested doses, treatment timings, responses from different cell types, or drug combinations. Further complexities include disentangling perturbation-specific effects from confounding variations, and modeling drug–drug interactions across diverse contexts may be challenging. In this section, we discuss recent exciting advances in ML methods that can be used to uncover biology from single-cell chemical genomics screens, as well as forward-looking directions involving pre-trained single-cell foundation models and single-cell causal discovery methods (Figure 3).

Modern representation learning methods have emerged as the computational backbone for single-cell chemical genomics analysis, leveraging the flexibility of variational autoencoders (VAEs) and hierarchical Bayesian modeling [117,118]. These approaches excel at reconstructing expression data while learning latent variables relating to cell states or modules of gene expression, providing denoising capabilities through probabilistic modeling [119–126]. Specifically, representation learning models feature generative processes that reconstruct expression profiles of seen conditions using approximated probability distributions of latent factors to summarize cell state (encoders) and mappings from latent factor reduced dimensions to expression space (decoders). Notably, the decoding process can be leveraged to predict transcriptional treatment effects from counterfactual conditions not represented in the original training set, depending on model design [117].
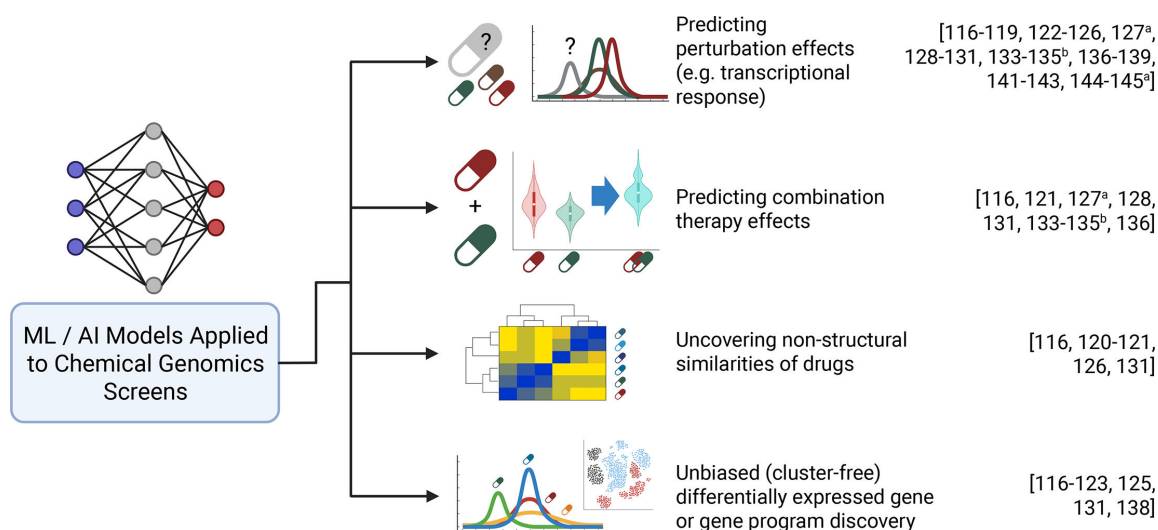

Figure 3: Overview of downstream analysis tasks performed by computational methods trained on chemical genomics screens.

Left, graphical representation and description of the task; right, citation number of methods that can perform the task. [a]Task achieved by method trained on genetic but not chemical perturbations as of publication. [b]Task possible through *in silico* gene regulatory network (GRN) perturbation but not explicitly tested by the method as of publication.

In just the past 2 years, an assortment of representational learning models has aimed to predict treatment effects on the single-cell transcriptome as a central task. The Compositional Perturbation Autoencoder (CPA) encodes conditions as independent covariates using adversarial loss to predict counterfactual expression profiles using linear combinations of latent factors and covariates [127]. This approach builds on concepts from the authors' prior scGen model that combines VAEs with latent space vector arithmetic to model how a perturbation shifts gene expression in unseen cell types [128] as well as their trVAE model that uses maximum mean discrepancy regularization to force cross-condition similarities and enable out-of-distribution predictions [129]. scVIDR additionally uses log-linear interpolation to predict cell type-specific, dose-dependent gene expression changes across unseen doses [130]. While these VAEs learn and encode perturbational effects as latent factors assumed to be consistent across cell states, MrVI (Multi-resolution Variational Inference) more generally models heterogeneous transcriptional response across nested multi-condition experiments, which can then be used for multiple downstream analyses, including computing transcriptomic similarities between seen and unseen conditions and predicting counterfactual differential expression for a given cell state [131]. Relatedly, DRVI employs additive decoders along with pooling functions to learn disentangled latent representations, in this case corresponding to distinct drug or gene perturbation transcriptional effects [132]. The sVAE+ and FCR methods aim to learn causal representations of chemical or genetic perturbations on single-cell transcriptomics readouts using sparse mechanism shift modeling and factorized causal representation learning, respectively [133,134]. Notably, FCR aims to distinguish treatment and covariate (e.g. cell line) specific effects, as well as interaction effects between treatments and other covariates, which may be useful for disease-specific prioritization of perturbation targets.

The recent scCADE method also tackles decoupling cellular context from perturbation responses using contrastive learning and attention mechanisms towards predicting unseen perturbation effects in seen cellular contexts [135]. Another recent method, LEMUR, avoids discrete categorization of cell states across experiment conditions and instead relies on matrix factorization using pre-specified conditions as covariates to map cell phenotypes to a continuous latent space and predict cell-state specific differential expression between conditions, although this approach may not reflect non-linear treatment effects [136]. OntoVAE incorporates biological ontologies into latent space representations, which allow for direct interpretation of simulated drug or gene transcriptional effects, although the model may not distinguish causal from direct relationships within the biological ontology space [137]. Conversely, scCAPE employs adversarial learning to distinguish perturbation effects on the single-cell transcriptome from intrinsic or extrinsic cell state variation, the latter being captured in scCAPE's latent space. However, scCAPE was originally designed for genetic perturbations (i.e. Perturb-seq training data), and perturbation inference is conducted at a factor and not a gene-specific level [138].

Modeling perturbations can be formulated as identifying a mapping between two distinct, unpaired single-cell expression datasets. As such, optimal transport has been proposed as a framework to infer or predict treatment effects. CINEMA-OT proposes optimal transport within a causal framework to isolate confounding variation and infer causal treatment effects at the single-cell level via counterfactual cell pairs, which can also be used to predict combinatorially perturbed phenotypes [139]. CellOT and the more high-dimensionally efficient W1-OT framework operate within a non-causal context to predict perturbation effects in single-cell expression data [140,141].

Finally, the BATCHIE framework utilizes active learning and flexible Bayesian modeling to enable adaptive experimental design for efficient and informative combinatorial drug screens. While not a fixed model, BATCHIE explicitly encourages iterative and rational drug screening to explore the combinatorial space, which we predict will be a crucial next direction in the field of chemical genomic screening [142]. In turn, rational generation of combinatorial perturbation data will further improve the performance of the aforementioned methods in modeling the full combinatorial perturbation space [127,132,138,139]. In summary, ML methods have shown great performance and promise in dissecting biology from large-scale experimental single-cell datasets.

Beyond predicting phenotypic responses, computational methods and causal discovery techniques are harnessing perturbation screens to reverse-engineer gene regulatory networks (GRNs). GRNs capture co-regulatory dynamics, specifically transcription factor (TF) and target gene interactions, and yield insight into the molecular underpinnings of disease. A well-constructed GRN could simulate dose-dependent or combinatorial effects of chemical or genetic perturbations by propagating perturbation effects across connected nodes (e.g. *in silico* knockdown, knockout, and/or overexpression of multiple TFs or kinases). Robust GRN inference from scRNA-seq has been a long-standing goal in the field of bioinformatics and

systems biology [143]. Unfortunately, directly inferring GRNs from experimental scRNA-seq alone, for example, through simplistic covariance analysis or specialized tools like SCENIC [144], often fails due to the static nature of single-timepoint gene expression and other technical limitations like noise and sparsity. An active area of interdisciplinary ML research applies novel causal discovery methods to better infer GRNs from genetic perturbation scRNA-seq [145,146]. Genetic perturbation scRNA-seq (e.g. Perturb-seq) is the gold-standard modality to study gene–gene relationships for GRN inference, although noise and sparsity inherent to genetic screens continue to impede robust GRN inference. Nonetheless, we anticipate future causal approaches will increasingly use deeper and wider chemical screens as training datasets instead.

Contemporary deep generative methods like PDGrapher, graphVCI, and TFdisc train on chemical, genetic, and/or no perturbation (wildtype, control) scRNA-seq datasets to predict transcriptional effects using already-constructed or inferred GRNs. PDGrapher predicts combination chemical perturbations necessary to reverse a given disease's effects given ground-truth GRNs, while TFdisc simulates TF perturbation by learning inter-regulation between TFs and target genes from wildtype scRNA-seq using random forests. graphVCI builds upon a previous variational causal inference framework VCI and trains on either chemical or genetic perturbation data to refine given ground-truth GRNs and predict cell-specific counterfactual transcriptional effects from unseen perturbations [147–150]. Key challenges remain in inferring GRNs, such as deconstructing cyclic relationships between feedback-looping genes in favor of directed acyclic graphs, but we are excited about the utility of these methods and their emerging application to analyzing chemical genomics screens.

Finally, in the advent of popular large language models enabled by generative pre-trained transformers (GPTs) such as ChatGPT, LLaMA, and DALL-E, we anticipate self-supervised foundation models (FMs) will play a large role in future efforts to predict drug-induced molecular phenotypes. FMs can be pre-trained on diverse, large-scale single-cell datasets to encode biology or other useful representations and are an attractive option in order to subsequently perform various downstream tasks with minimal-to-no refinement. Several single-cell FMs have been developed with diverse performance across common analysis tasks within single-cell genomics pipelines, with impressive performance already noted in automated cell type annotation and strong headways in predicting gene expression changes following genetic perturbations [151]. Notably, scGPT can predict unseen genetic perturbation responses and infer genetic networks, scFoundation features both drug and genetic perturbation response predictions as downstream tasks, and other single-cell FMs may indirectly yield insight into drug response at the expression level across tissue types or species [152–156]. While FMs already show promise in single-cell data analysis, their application to chemical genomic screens to perform downstream analyses such as predicting dose-dependent or combinatorial drug effects is yet to be tested, and it remains unclear whether they currently outperform the aforementioned methods.

We anticipate a new era of transfer learning in chemical genomics where pre-trained models can be adapted to predict any chemical or genetic perturbation combination in diverse contexts. Such *in silico* predictions could greatly accelerate virtual screening of drug effects, hypothesis generation for combination therapies, and discovery of causal regulatory mechanisms through providing a flexible backbone that had already learned the language of cells [157]. A central challenge is the lack of a large corpus of single-cell chemical genomics datasets necessary for pre-training existing and future model architectures and refining single-cell FM pre-training strategies. While biopharma industry-sponsored initiatives have made headway into generating large-scale, single-cell drug perturbation datasets [75], we predict comprehensive training datasets in single-cell genomics will remain elusive in the near term. Future efforts in integrating causal structures into attention-based FMs would be exciting directions to ensure interpretability.

Outlook and future perspectives

The integration of single-cell molecular profiling into chemical genomics and drug discovery has demonstrated its potential to transform our understanding of drug mechanisms, polypharmacology, and therapeutic efficacy. Looking ahead, we expect that advancements in single-cell methodologies will greatly refine our ability to perform drug target identification, characterize correlates (biomarkers) of therapeutic response, and predict resistance mechanisms a priori.

The resolution provided by single-cell genomics has provided key insights into the cell type(s) most likely to be affected in a disease and make predictions regarding how cellular heterogeneity and shifts in cell state(s) co-vary with disease phenotypes and an individual's response to therapy [15,17,62,158–163]. This

ability to refine target identification can help ensure that therapeutic interventions are directed at the most relevant cellular subpopulations. In the context of the chemical genetics paradigm, single-cell chemical genomics profiling has more deeply defined the consequence of drug exposure capturing molecular and cellular determinants of efficacy and resistance [62,114,164,165], which can ultimately guide the design of more effective combination therapies.

In the context of cancer, various screenable modeling approaches have been developed to facilitate the high-throughput screening of drug response at single-cell resolution. Established cell lines using multiplexed sequencing methods have provided insights into how genetic background influences drug response [5,102] and are rapidly scalable and multiplexable [75]. Patient-derived models have been optimized to maintain tumor heterogeneity [100,166], which can improve the clinical relevance of preclinical findings. Organoid models have allowed for the study of drug responses in three-dimensional environments that mimic *in vivo* conditions [167], while explants have preserved native tissue architecture and cell–cell interactions [165,168]. However, a systematic analysis to identify minimal screenable systems for a given molecular phenotype (i.e. retain key disease and therapeutic response features while maintaining multiplexing ability) is still lacking. Looking ahead, extending single-cell chemical genomics to more physiologically relevant model systems will be essential. Organoids [169], patient-derived explants [165], and microphysiological 'organ-on-chip' platforms [170] can better capture native 3D architecture, microenvironment interactions, and gradients in drug exposure that shape therapeutic response. Key opportunities lie in adapting these complex systems to high-throughput profiling and perturbation screens and in integrating spatially resolved readouts that map transcriptional drug responses back to defined structural features. Lastly, evolution under therapy due to genetic variation [171] or cellular plasticity [172] has also been identified as a manner by which tumors evade the effects of therapy. Incorporation of single-cell lineage tracing [104,173] will be instrumental in determining how cell populations evolve in response to drug exposure and has already highlighted how clonal expansions contribute to therapeutic resistance to chemotherapy [174].

Future promising directions in single-cell chemical genomics include the integration of spatial transcriptomics [175–179] and spatially barcoded single-cell transcriptomics approaches [180,181] to capture tissue-specific drug responses to drugs and leverage optical molecular screens [182–184] to define their regulation. While artificial intelligence (AI) and ML hold promise for advancing drug discovery, a major challenge lies in the need for comprehensive, large-scale, high-quality single-cell perturbation datasets to train these models effectively, a challenge that is excitingly beginning to be met by the release of the Tahoe-100M datasets [75].

The next decade holds immense promise for single-cell molecular profiling in drug discovery. By integrating high-throughput single-cell screening with computational innovations and multi-modal approaches, researchers will unlock deeper insights into drug mechanisms of action, accelerate the identification of effective therapies, and refine personalized treatment strategies. While challenges remain in scalability, data interpretation, and clinical translation, continued interdisciplinary advancements will solidify single-cell chemical genomics as a cornerstone of modern drug development.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Open Access

Open access for this article was enabled by the participation of Columbia University in an all-inclusive Read & Publish agreement with Portland Press and the Biochemical Society.

CRediT Author Contribution

A.W. contributed to writing—original draft and writing—review and editing. K.H.-H. contributed to writing—original draft and writing—review and editing. R.M.G. contributed to writing—original draft and writing—review and editing. E.A. contributed to supervision and writing—review and editing. J.L.M.-F. contributed to writing—original draft and writing—review and editing.

Acknowledgments

We thank the members of the McFaline-Figueroa and Azizi labs for helpful comments and suggestions in the preparation of this manuscript.

Abbreviations

ATAC-seq, assay for transposase accessible chromatin; CMap, Connectivity Map; FMs, foundation models; GPTs, generative pre-trained transformers; GRN, gene regulatory network; ML, machine learning; NGS, next-generation sequencing; PDD, phenotypic drug discovery; STAMPs, single-cell transcriptomes attached to microparticles; TDD, target-based drug discovery; TF, transcription factor; VAE, variational autoencoders; sgRNA, single guide RNA.

References

1   Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., et al. (2017) Defining a cancer dependency map. *Cell* **170**, 564–576 https://doi.org/10.1016/j.cell.2017.06.010 PMID: 28753430

2   Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X, et al. (2017) A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452 https://doi.org/10.1016/j.cell.2017.10.049 PMID: 29195078

3   Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J. et al. (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **4**, 1293–1301 https://doi.org/10.1038/3282 PMID: 9809554

4   Bush, E.C., Ray, F., Alvarez, M.J., Realubit, R., Li, H., Karan, C. et al. (2017) PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nat. Commun.* **8**, 105 https://doi.org/10.1038/s41467-017-00136-z PMID: 28740083

5   Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J., Packer, J. et al. (2020) Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45–51 https://doi.org/10.1126/science.aax6234 PMID: 31806696

6   Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Guna, A, et al. (2022) Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575 https://doi.org/10.1016/j.cell.2022.05.013 PMID: 35688146

7   Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E, et al. (2018) Cell-of-origin patterns dominate the molecular classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304 https://doi.org/10.1016/j.cell.2018.03.022 PMID: 29625048

8   Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M. et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 https://doi.org/10.1038/ng.499 PMID: 19915526

9   Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N. et al. (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 https://doi.org/10.1038/s41588-022-01024-z PMID: 35379992

10 Fowler, D.M. and Fields, S. (2014) Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 https://doi.org/10.1038/nmeth.3027 PMID: 25075907

11 Bock, C., Datlinger, P., Chardon, F., Coelho, M.A., Dong, M.B., Lawson, K.A. et al. (2022) High-content CRISPR screening. *Nat. Rev. Methods Primers* **2** https://doi.org/10.1038/s43586-021-00093-4

12 Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 https://doi.org/10.1126/science.1246981 PMID: 24336569

13 Emani, P.S., Liu, J.J., Clarke, D., Jensen, M., Warrell, J., Gupta, C. et al. (2024) Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 https://doi.org/10.1126/science.adi5199 PMID: 38781369

14 Park, J.H., Hothi, P., De Lomana, A.L.G., Pan, M., Calder, R., Turkarslan, S, et al. (2024) Gene regulatory network topology governs resistance and treatment escape in glioma stem-like cells. *Sci. Adv* **10**, eadj7706 https://doi.org/10.1126/sciadv.adj7706 PMID: 38848360

15 Kunes, R.Z., Walle, T., Land, M., Nawy, T. and Pe'er, D. (2024) Supervised discovery of interpretable gene programs from single-cell data. *Nat. Biotechnol.* **42**, 1084–1095 https://doi.org/10.1038/s41587-023-01940-3 PMID: 37735262

16 Weinstock, J.S., Arce, M.M., Freimer, J.W., Ota, M., Marson, A., Battle, A. et al. (2024) Gene regulatory network inference from CRISPR perturbations in primary CD4+ T cells elucidates the genomic basis of immune disease. *Cell Genom.* **4**, 100671 https://doi.org/10.1016/j.xgen.2024.100671 PMID: 39395408

17 Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E. et al. (2018) A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature New Biol.* **560**, 319–324 https://doi.org/10.1038/s41586-018-0393-7 PMID: 30069044

18 He, S., Jin, Y., Nazaret, A., Shi, L., Chen, X., Rampersaud, S. et al. (2025) Starfysh integrates spatial transcriptomic and histologic data to reveal heterogeneous tumor-immune hubs. *Nat. Biotechnol.* **43**, 223–235 https://doi.org/10.1038/s41587-024-02173-8 PMID: 38514799

19 Stockwell, B.R. (2000) Chemical genetics: ligand-based discovery of gene function. *Nat. Rev. Genet.* **1**, 116–125 https://doi.org/10.1038/35038557 PMID: 11253651

20   Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T. et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 https://doi.org/10.1126/science.1247005 PMID:24336571

21   MacBeath, G. (2001) Chemical genomics: what will it take and who gets to play? *Genome Biol.* **2**, COMMENT2005 https://doi.org/10.1186/gb-2001-2-6-comment2005 PMID:11423004

22   Swinney, D.C. (2013) Phenotypic vs. target-based drug discovery for first-in-class medicines. *Clin. Pharmacol. Ther.* **93**, 299–301 https://doi.org/10.1038/clpt.2012.236 PMID:23511784

23   Pun, F.W., Ozerov, I.V. and Zhavoronkov, A. (2023) AI-powered therapeutic target discovery. *Trends Pharmacol. Sci* **44**, 561–572 https://doi.org/10.1016/j.tips.2023.06.010 PMID:37479540

24   Vincent, F., Nueda, A., Lee, J., Schenone, M., Prunotto, M. and Mercola, M. (2022) Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nat. Rev. Drug Discov.* **21**, 899–914 https://doi.org/10.1038/s41573-022-00472-w PMID:35637317

25   Kumar, V., Chunchagatta Lakshman, P.K., Prasad, T.K., Manjunath, K., Bairy, S., Vasu, A.S. et al. (2024) Target-based drug discovery: applications of fluorescence techniques in high throughput and fragment-based screening. *Heliyon* **10**, e23864 https://doi.org/10.1016/j.heliyon.2023.e23864 PMID:38226204

26   Lin, A., Giuliano, C.J., Palladino, A., John, K.M., Abramowicz, C., Yuan, M.L. et al. (2019) Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11**, eaaw8412 https://doi.org/10.1126/scitranslmed.aaw8412 PMID:31511426

27   Sun, D., Gao, W., Hu, H. and Zhou, S. (2022) Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* **12**, 3049–3062 https://doi.org/10.1016/j.apsb.2022.02.002 PMID:35865092

28   Harrison, R.K. (2016) Phase II and phase III failures: 2013-2015. *Nat. Rev. Drug Discov.* **15**, 817–818 https://doi.org/10.1038/nrd.2016.184 PMID:27811931

29   Van de Sande, B., Lee, J.S., Mutasa-Gottgens, E., Naughton, B., Bacon, W., Manning, J. et al. (2023) Applications of single-cell RNA sequencing in drug discovery and development. *Nat. Rev. Drug Discov.* **22**, 496–520 https://doi.org/10.1038/s41573-023-00688-4 PMID:37117846

30   Lombardino, J.G. and Lowe, J.A., III. (2004) The role of the medicinal chemist in drug discovery--then and now. *Nat. Rev. Drug Discov.* **3**, 853–862 https://doi.org/10.1038/nrd1523 PMID:15459676

31   Cuomo, A.S.E., Nathan, A., Raychaudhuri, S., MacArthur, D.G. and Powell, J.E. (2023) Single-cell genomics meets human genetics. *Nat. Rev. Genet.* **24**, 535–549 https://doi.org/10.1038/s41576-023-00599-5 PMID:37085594

32   Luecken, M.D. and Theis, F.J. (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 https://doi.org/10.15252/msb.20188746 PMID:31217225

33   De Jonghe, J., Opzoomer, J.W., Vilas-Zornoza, A., Nilges, B.S., Crane, P., Vicari, M. et al. (2024) scTrends: A living review of commercial single-cell and spatial 'omic technologies. *Cell Genom.* **4**, 100723 https://doi.org/10.1016/j.xgen.2024.100723 PMID:39667347

34   Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 https://doi.org/10.1126/science.270.5235.467 PMID:7569999

35   Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D. et al. (2000) Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 https://doi.org/10.1016/s0092-8674(00)00015-5 PMID:10929718

36   Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D, et al. (2018) The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. *Cell Syst* **6**, 13–24 https://doi.org/10.1016/j.cels.2017.11.001 PMID:29199020

37   Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 https://doi.org/10.1126/science.1132939 PMID:17008526

38   Library of integrated network-based cellular signatures (LINCS). Virtual symposium. https://commonfund.nih.gov/sites/default/files/LINCS_Final_Symposium_Summary_Final_508.pdf

39   Vazquez, F. and Boehm, J.S. (2020) The cancer dependency map enables drug mechanism-of-action investigations. *Mol. Syst. Biol.* **16**, e9757 https://doi.org/10.15252/msb.20209757 PMID:32696566

40   Arafeh, R., Shibue, T., Dempster, J.M., Hahn, W.C. and Vazquez, F. (2025) The present and future of the Cancer Dependency Map. *Nat. Rev. Cancer* **25**, 59–73 https://doi.org/10.1038/s41568-024-00763-x PMID:39468210

41   Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., III et al. (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 https://doi.org/10.1038/s41586-019-1186-3

42   Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S. et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 https://doi.org/10.1038/nature11003 PMID:22460905

43   Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S. et al. (2016) Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 https://doi.org/10.1038/nchembio.1986 PMID:26656090

44   Yu, C., Mannan, A.M., Yvone, G.M., Ross, K.N., Zhang, Y.-L., Marton, M.A. et al. (2016) High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 https://doi.org/10.1038/nbt.3460 PMID:26928769

45   Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E. et al. (2017) The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* **23**, 405–408 https://doi.org/10.1038/nm.4306 PMID:28388612

46   Ye, C., Ho, D.J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R. et al. (2018) DRUG-seq for miniaturized high-throughput transcriptome profiling in DRUG discovery. *Nat. Commun.* **9**, 4307 https://doi.org/10.1038/s41467-018-06500-x PMID:30333485

47   Moellering, R.E. and Cravatt, B.F. (2012) How chemoproteomics can enable drug discovery and development. *Chem. Biol.* **19**, 11–22 https://doi.org/10.1016/j.chembiol.2012.01.001 PMID:22284350

48   Bantscheff, M., Hopf, C., Savitski, M.M., Dittmann, A., Grandi, P., Michon, A.-M. et al. (2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* **29**, 255–265 https://doi.org/10.1038/nbt.1759 PMID:21258344

49   Lin, C., Sniezek, C.M., McGann, C.D., Karki, R., Giglio, R.M., Garcia, B.A. et al. (2024) Defining the heterogeneous molecular landscape of lung cancer cell responses to epigenetic inhibition. *bioRxiv*. https://doi.org/10.1101/2024.05.23.592075

50   Litichevskiy, L., Peckner, R., Abelin, J.G., Asiedu, J.K., Creech, A.L., Davis, J.F, et al. (2018) A Library of Phosphoproteomic and Chromatin Signatures for Characterizing Cellular Responses to Drug Perturbations. *Cell Syst* **6**, 424–443 https://doi.org/10.1016/j.cels.2018.03.012 PMID: 29655704

51   Mitchell, D.C., Kuljanin, M., Li, J., Van Vranken, J.G., Bulloch, N., Schweppe, D.K. et al. (2023) A proteome-wide atlas of drug mechanism of action. *Nat. Biotechnol.* **41**, 845–857 https://doi.org/10.1038/s41587-022-01539-0 PMID: 36593396

52   Vranken, J.G., Li, J., Mintseris, J., Gadzuk-Shea, M., Gygi, S.P. and Schweppe, D.K. (2024) Large-scale characterization of drug mechanism of action using proteome-wide thermal shift assays. *bioRxiv* https://doi.org/10.1101/2024.01.26.577428

53   Flower, C.T., Liu, C., Chuang, H.-Y., Ye, X., Cheng, H., Heath, J.R, et al. (2024) Signaling and transcriptional dynamics underlying early adaptation to oncogenic BRAF inhibition. *bioRxiv* https://doi.org/10.1101/2024.02.19.581004

54   Zi, Z., Zhang, Y., Zhang, P., Ding, Q., Chu, M., Chen, Y. et al. (2020) A proteomic connectivity map for characterizing the tumor adaptive response to small molecule chemical perturbagens. *ACS Chem. Biol.* **15**, 140–150 https://doi.org/10.1021/acschembio.9b00694 PMID: 31846293

55   Ben-David, U., Siranosian, B., Ha, G., Tang, H., Oren, Y., Hinohara, K. et al. (2018) Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325–330 https://doi.org/10.1038/s41586-018-0409-3 PMID: 30089904

56   Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N. et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 https://doi.org/10.1038/nmeth.1315 PMID: 19349980

57   Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H. et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 https://doi.org/10.1126/science.1254257 PMID: 24925914

58   McGranahan, N. and Swanton, C. (2015) Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 https://doi.org/10.1016/j.ccell.2014.12.001 PMID: 25584892

59   McFaline-Figueroa, J.L., Hill, A.J., Qiu, X., Jackson, D., Shendure, J. and Trapnell, C. (2019) A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat. Genet.* **51**, 1389–1398 https://doi.org/10.1038/s41588-019-0489-5 PMID: 31477929

60   Kinker, G.S., Greenwald, A.C., Tal, R., Orlova, Z., Cuoco, M.S., McFarland, J.M. et al. (2020) Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 https://doi.org/10.1038/s41588-020-00726-6 PMID: 33128048

61   Sprenger, J. and Weinberger, N. (2021) Simpson's Paradox. The Stanford Encyclopedia of Philosophy. In *Metaphysics Research Lab* (E.N. Zalta, ed), Stanford University

62   França, G.S., Baron, M., King, B.R., Bossowski, J.P., Bjornberg, A., Pour, M. et al. (2024) Cellular adaptation to cancer therapy along a resistance continuum. *Nature* **631**, 876–883 https://doi.org/10.1038/s41586-024-07690-9

63   Kim, D.W., Tu, K.J., Wei, A., Lau, A.J., Gonzalez-Gil, A., Cao, T. et al. (2022) Amyloid-beta and tau pathologies act synergistically to induce novel disease stage-specific microglia subtypes. *Mol. Neurodegener.* **17**, 83 https://doi.org/10.1186/s13024-022-00589-x

64   Grandke, F., Fehlmann, T., Kern, F., Gate, D.M., Wolff, T.W., Leventhal, O. et al. (2025) A single-cell atlas to map sex-specific gene-expression changes in blood upon neurodegeneration. *Nat. Commun.* **16** https://doi.org/10.1038/s41467-025-56833-7

65   Gierahn, T.M., Wadsworth, M.H., II, Hughes, T.K., Bryson, B.D., Butler, A., Satija, R. et al. (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 https://doi.org/10.1038/nmeth.4179

66   Goldstein, L.D., Chen, Y.-J.J., Dunne, J., Mir, A., Hubschle, H., Guillory, J. et al. (2017) Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics* **18**, 519 https://doi.org/10.1186/s12864-017-3893-1 PMID: 28687070

67   Prakadan, S.M., Shalek, A.K. and Weitz, D.A. (2017) Scaling by shrinking: empowering single-cell "omics" with microfluidic devices. *Nat. Rev. Genet.* **18**, 345–361 https://doi.org/10.1038/nrg.2017.15 PMID: 28392571

68   Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M, et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 https://doi.org/10.1016/j.cell.2015.05.002 PMID: 26000488

69   Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V, et al. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 https://doi.org/10.1016/j.cell.2015.04.044 PMID: 26000487

70   Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R. et al. (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 https://doi.org/10.1126/science.aam8940 PMID: 28818938

71   Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z. et al. (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 https://doi.org/10.1126/science.aam8999 PMID: 29545511

72   Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J. et al. (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 https://doi.org/10.1038/s41586-019-0969-x

73   Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H. et al. (2019) A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science* **365**, eaax1971 https://doi.org/10.1126/science.aax1971 PMID: 31488706

74   Svensson, V., Vento-Tormo, R. and Teichmann, S.A. (2018) Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 https://doi.org/10.1038/nprot.2017.149

75   Zhang, J., Ubas, A.A., De Borja, R., Svensson, V., Thomas, N., Thakar, N, et al. (2025) Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. *Genomics* https://doi.org/10.1101/2025.02.20.639398

76   Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J. et al. (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 https://doi.org/10.1038/nmeth.3370

77   Zhu, C., Yu, M., Huang, H., Juric, I., Abnousi, A., Hu, R. et al. (2019) An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 https://doi.org/10.1038/s41594-019-0323-x PMID: 31695190

78   Reuter, J.A., Spacek, D.V., Pai, R.K. and Snyder, M.P. (2016) Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat. Methods* **13**, 953–958 https://doi.org/10.1038/nmeth.4028 PMID: 27723755

79   Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C. et al. (2018) scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 https://doi.org/10.1038/s41467-018-03149-4 PMID: 29472610

80   Swanson, E., Lord, C., Reading, J., Heubeck, A.T., Genge, P.C., Thomson, Z. et al. (2021) Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife* **10**, e63632 https://doi.org/10.7554/eLife.63632 PMID: 33835024

81 Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C. et al. (2017) Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 https://doi.org/10.1038/nbt.3973

82 Yuan, J., Sheng, J. and Sims, P.A. (2018) SCOPE-Seq: a scalable technology for linking live cell imaging and single-cell RNA sequencing. *Genome Biol.* **19**, 227 https://doi.org/10.1186/s13059-018-1607-x PMID: 30583733

83 Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K. and Swerdlow, H. (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 https://doi.org/10.1038/nmeth.4380

84 Blair, J.D., Hartman, A., Zenk, F., Wahle, P., Brancati, G., Dalgarno, C. et al. (2025) Phospho-seq: integrated, multi-modal profiling of intracellular protein dynamics in single cells. *Nat. Commun.* **16**, 1346 https://doi.org/10.1038/s41467-025-56590-7 PMID: 39905064

85 Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 https://doi.org/10.1038/nmeth.2688

86 Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P. et al. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 https://doi.org/10.1038/nature14590 PMID: 26083756

87 Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L. et al. (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 https://doi.org/10.1126/science.aab1601 PMID: 25953818

88 Fiskin, E., Lareau, C.A., Ludwig, L.S., Eraslan, G., Liu, F., Ring, A.M. et al. (2022) Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat. Biotechnol.* **40**, 374–381 https://doi.org/10.1038/s41587-021-01065-5 PMID: 34675424

89 Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y, et al. (2020) Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103–1116 https://doi.org/10.1016/j.cell.2020.09.056 PMID: 33098772

90 Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J. et al. (2018) Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 https://doi.org/10.1126/science.aau0730 PMID: 30166440

91 Olsen, T.R., Talla, P., Sagatelian, R.K., Furnari, J., Bruce, J.N., Canoll, P. et al. (2025) Scalable co-sequencing of RNA and DNA from individual nuclei. *Nat. Methods* **22**, 477–487 https://doi.org/10.1038/s41592-024-02579-x PMID: 39939719

92 Chen, A.F., Parks, B., Kathiria, A.S., Ober-Reynolds, B., Goronzy, J.J. and Greenleaf, W.J. (2022) NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 https://doi.org/10.1038/s41592-022-01461-y PMID: 35501385

93 Nichols, R.V., O'Connell, B.L., Mulqueen, R.M., Thomas, J., Woodfin, A.R., Acharya, S. et al. (2022) High-throughput robust single-cell DNA methylation profiling with sciMETv2. *Nat. Commun.* **13**, 7627 https://doi.org/10.1038/s41467-022-35374-3 PMID: 36494343

94 Fan, X., Yang, C., Li, W., Bai, X., Zhou, X., Xie, H. et al. (2021) SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol.* **22**, 195 https://doi.org/10.1186/s13059-021-02406-y PMID: 34193232

95 Hartmann, F.J., Mrdjen, D., McCaffrey, E., Glass, D.R., Greenwald, N.F., Bharadwaj, A. et al. (2021) Single-cell metabolic profiling of human cytotoxic T cells. *Nat. Biotechnol.* **39**, 186–197 https://doi.org/10.1038/s41587-020-0651-8 PMID: 32868913

96 Lopez, T., Honigfort, D., Mah, A., Thompson, C., Dien, V., Dawood, M, et al. High-Throughput Multiomics Profiling of Model Systems Using the AVITI24 Platform. *Genomics*. https://doi.org/10.1101/2025.05.03.651997

97 What is Sample Multiplexing? Official 10x Genomics Support. https://www.10xgenomics.com/support/software/cell-ranger/latest/getting-started/cr-3p-what-is-cellplex

98 Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., 3rd et al. (2018) Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 https://doi.org/10.1186/s13059-018-1603-1 PMID: 30567574

99 McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V. et al. (2019) MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 https://doi.org/10.1038/s41592-019-0433-8 PMID: 31209384

100 Giglio, R.M., Hou, N., Wyatt, A., Hong, J., Shi, L., Vaikunthan, M, et al. (2025) A heterogeneous pharmaco-transcriptomic landscape induced by targeting a single oncogenic kinase. 2024.04.08.587960 bioRxiv. https://doi.org/10.1101/2024.04.08.587960 PMID: 38645018

101 Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E. et al. (2018) Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 https://doi.org/10.1038/nbt.4042 PMID: 29227470

102 McFarland, J.M., Paolella, B.R., Warren, A., Geiger-Schuller, K., Shibue, T., Rothberg, M. et al. (2020) Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat. Commun.* **11**, 4296, 4296 https://doi.org/10.1038/s41467-020-17440-w PMID: 32855387

103 Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T. et al. (2018) Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 https://doi.org/10.1038/s41586-018-0744-4 PMID: 30518857

104 Jindal, K., Adil, M.T., Yamaguchi, N., Yang, X., Wang, H.C., Kamimoto, K. et al. (2024) Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. *Nat. Biotechnol.* **42**, 946–959 https://doi.org/10.1038/s41587-023-01931-4 PMID: 37749269

105 Cheng, Y.L., Banu, M.A., Zhao, W., Rosenfeld, S.S., Canoll, P. and Sims, P.A. (2023) Multiplexed single-cell lineage tracing of mitotic kinesin inhibitor resistance in glioblastoma. *Cancer Biology*. Sep 12, 2023. https://doi.org/10.1101/2023.09.09.557001

106 Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L. et al. (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866. https://doi.org/10.1016/j.cell.2016.11.038

107 Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y, et al. (2016) A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882 https://doi.org/10.1016/j.cell.2016.11.048 PMID: 27984733

108 Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E. et al. (2016) Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883–1896. https://doi.org/10.1016/j.cell.2016.11.039

109 Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J. et al. (2017) Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 https://doi.org/10.1038/nmeth.4177

110 Xie, S., Duan, J., Li, B., Zhou, P. and Hon, G.C. (2017) Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* **66**, 285–299. https://doi.org/10.1016/j.molcel.2017.03.007

111 Hill, A.J., McFaline-Figueroa, J.L., Starita, L.M., Gasperini, M.J., Matreyek, K.A., Packer, J. et al. (2018) On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 https://doi.org/10.1038/nmeth.4604 PMID: 29457792

112 Replogle, J.M., Norman, T.M., Xu, A., Hussmann, J.A., Chen, J., Cogan, J.Z. et al. (2020) Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 https://doi.org/10.1038/s41587-020-0470-y PMID: 32231336

113 Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M. et al. (2019) Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 https://doi.org/10.1038/s41592-019-0392-0 PMID: 31011186

114 McFaline-Figueroa, J.L., Srivatsan, S., Hill, A.J., Gasperini, M., Jackson, D.L., Saunders, L. et al. (2024) Multiplex single-cell chemical genomics reveals the kinase dependence of the response to targeted therapy. *Cell Genom.* **4**, 100487 https://doi.org/10.1016/j.xgen.2023.100487 PMID: 38278156

115 Xu, Z., Sziraki, A., Lee, J., Zhou, W. and Cao, J. (2024) Dissecting key regulators of transcriptome kinetics through scalable single-cell RNA profiling of pooled CRISPR screens. *Nat. Biotechnol.* **42**, 1218–1223 https://doi.org/10.1038/s41587-023-01948-9 PMID: 37749268

116 Ji, Y., Lotfollahi, M., Wolf, F.A. and Theis, F.J. (2021) Machine learning for perturbational single-cell omics. *Cell Syst.* **12**, 522–537 https://doi.org/10.1016/j.cels.2021.05.016 PMID: 34139164

117 Kingma, D.P. and Welling, M. (2013) *Auto-Encoding Variational Bayes*

118 Gopalan, P., Hofman, J.M. and Blei, D.M. (2013) Scalable recommendation with Poisson factorization. *arXiv [cs.IR]* https://doi.org/10.48550/ARXIV.1311.1704

119 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 https://doi.org/10.1038/s41592-018-0229-2 PMID: 30504886

120 Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J. et al. (2022) A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 https://doi.org/10.1038/s41587-021-01206-w PMID: 35132262

121 Wagner, A., Regev, A. and Yosef, N. (2016) Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160 https://doi.org/10.1038/nbt.3711 PMID: 27824854

122 Levitin, H.M., Yuan, J., Cheng, Y.L., Ruiz, F.J., Bush, E.C., Bruce, J.N. et al. (2019) *De novo* gene signature identification from single-cell RNA-seq with hierarchical Poisson factorization. *Mol. Syst. Biol.* **15**, e8557 https://doi.org/10.15252/msb.20188557 PMID: 30796088

123 Prabhakaran, S., Azizi, E., Carr, A. and Pe'er, D. (2016) Dirichlet Process Mixture Model for Correcting Technical Variation in Single-Cell Gene Expression Data. In International Conference on Machine Learning, pp. 1070–1079, PMLR

124 Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., et al. (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 https://doi.org/10.1186/s13059-015-0844-5 PMID: 26653891

125 Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S. and Theis, F.J. (2019) Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10**, 1–14 https://doi.org/10.1038/s41467-018-07931-2

126 Grønbech, C.H., Vording, M.F., Timshel, P.N., Sønderby, C.K., Pers, T.H. and Winther, O. (2020) scVAE: variational auto-encoders for single-cell gene expression data. *Bioinformatics* **36**, 4415–4422 https://doi.org/10.1093/bioinformatics/btaa293 PMID: 32415966

127 Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., Hetzel, L., Ji, Y., Ibarra, I.L. et al. (2023) Predicting cellular responses to complex perturbations in high-throughput screens. *Mol. Syst. Biol.* **19**, e11517 https://doi.org/10.15252/msb.202211517 PMID: 37154091

128 Lotfollahi, M., Wolf, F.A. and Theis, F.J. (2019) scGen predicts single-cell perturbation responses. *Nat. Methods* **16**, 715–721 https://doi.org/10.1038/s41592-019-0494-8 PMID: 31363220

129 Lotfollahi, M., Naghipourfar, M., Theis, F.J. and Wolf, F.A. (2020) Conditional out-of-distribution generation for unpaired data using transfer VAE. *Bioinformatics* **36**, i610–i617 https://doi.org/10.1093/bioinformatics/btaa800 PMID: 33381839

130 Kana, O., Nault, R., Filipovic, D., Marri, D., Zacharewski, T. and Bhattacharya, S. (2023) Generative modeling of single-cell gene expression for dose-dependent chemical perturbations. *Patterns (N Y)* **4**, 100817 https://doi.org/10.1016/j.patter.2023.100817 PMID: 37602218

131 Boyeau, P., Hong, J., Gayoso, A., Kim, M., McFaline-Figueroa, J.L., Jordan, M.I., et al. (2022) Deep generative modeling of sample-level heterogeneity in single-cell genomics. *Bioinformatics* https://doi.org/10.1101/2022.10.04.510898

132 Moinfar, A.A. and Theis, F.J. (2024) Unsupervised deep disentangled representation of single-cell omics. *bioRxiv.* https://doi.org/10.1101/2024.11.06.622266

133 Lopez, R., Tagasovska, N., Ra, S., Cho, K., Pritchard, J.K. and Regev, A. (2022) *Learning Causal Representations of Single Cells via Sparse Mechanism Shift Modeling*

134 Mao, H., Lopez, R., Liu, K., Hütter, J.-C., Richmond, D., Benos, P.V., et al. (2024) *Learning Identifiable Factorized Causal Representations of Cellular Responses*

135 Ou, J., Li, J., Xia, Z., Dai, S., Ding, Y., Guo, Y, et al. scCADE: A Superior Tool for Predicting Perturbation Responses in Single-Cell Gene Expression Using Contrastive Learning and Attention Mechanisms. https://ieeexplore.ieee.org/abstract/document/10822339?casa_token=MMvL1spuicYAAAAA:agvAseckFcO1fPKCoBY3fs9Rw0qy01zTJR2IeE3WmS0Xg86GnCuSVsdUEH7mZg958j4uqV3QnLdU

136 Ahlmann-Eltze, C. and Huber, W. (2025) Analysis of multi-condition single-cell data with latent embedding multivariate regression. **57**, 659–667 *bioRxiv.* https://doi.org/10.1101/2023.03.06.531268

137 Doncevic, D. and Herrmann, C. (2023) Biologically informed variational autoencoders allow predictive modeling of genetic and drug-induced perturbations. *Bioinformatics* **39**, btad387 https://doi.org/10.1093/bioinformatics/btad387 PMID: 37326971

138 Fu, Z. and Hou, L. (2024) Inference of heterogeneous effects in single-cell genetic perturbation screens. *bioRxiv.* https://doi.org/10.1101/2024.05.28.596224

139 Dong, M., Wang, B., Wei, J., Deo. Fonseca, A.H., Perry, C.J., Frey, A, et al. (2023) Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nat. Methods* **20**, 1769–1779 https://doi.org/10.1038/s41592-023-02040-5

140 Bunne, C., Stark, S.G., Gut, G., Del Castillo, J.S., Levesque, M., Lehmann, K.V., et al. (2023) Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* **20**, 1759–1768 https://doi.org/10.1038/s41592-023-01969-x

141 Chen, Y., Hu, Z., Chen, W. and Huang, H. (2025) Fast and scalable Wasserstein-1 neural optimal transport solver for single-cell perturbation prediction. *Bioinformatics* **41**, i513–i522 https://doi.org/10.1093/bioinformatics/btaf253

142 Tosh, C., Tec, M., White, J., Quinn, J.F., Ibanez Sanchez, G., Calder, P. et al. (2023) A Bayesian active learning platform for scalable combination drug screens. *bioRxiv*. https://doi.org/10.1101/2023.12.18.572245

143 Madhamshettiwar, P.B., Maetschke, S.R., Davis, M.J., Reverter, A. and Ragan, M.A. (2012) Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Med.* **4**, 41 https://doi.org/10.1186/gm340 PMID: 22548828

144 Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G. et al. (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 https://doi.org/10.1038/nmeth.4463 PMID: 28991892

145 Lopez, R., Hütter, J.C., Pritchard, J.K. and Regev, A. (2022) Large-scale Differentiable Causal Discovery of Factor Graphs. *arXiv* https://doi.org/10.48550/ARXIV.2206.07824

146 Nazaret, A., Hong, J., Azizi, E. and Blei, D. (2023) Stable Differentiable Causal Discovery. *arXiv* https://doi.org/10.48550/ARXIV.2311.10263

147 Gonzalez, G., Lin, X., Herath, I., Veselkov, K., Bronstein, M. and Zitnik, M. (2024) Combinatorial prediction of therapeutic perturbations using causally-inspired neural networks. *bioRxiv*. https://doi.org/10.1101/2024.01.03.573985

148 Barton, R., Wang, Z., Ioannidis, V.N., Donno, C., Price, L.C., et al. (2022) Predicting Cellular Responses with Variational Causal Inference and Refined Relational Information. In The Eleventh International Conference on Learning Representations

149 Wang, H., Li, Q. and Zhu, Y. (2024) A novel in silico transcription factor perturbation simulator uncovers diverse genetic architecture of brain disorders. *In Review*. https://doi.org/10.21203/rs.3.rs-5290734/v1

150 Wu, Y., Price, L.C., Wang, Z., Ioannidis, V.N., Barton, R.A. and Karypis, G. (2022) Variational Causal Inference. *arXiv [stat.ML]* https://doi.org/10.48550/ARXIV.2209.05935

151 Liu, T., Li, K., Wang, Y., Li, H. and Zhao, H. Evaluating the Utilities of Foundation Models in Single-cell Data Analysis. *Bioinformatics*. https://doi.org/10.1101/2023.09.08.555192

152 Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N. et al. (2024) scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 https://doi.org/10.1038/s41592-024-02201-0 PMID: 38409223

153 Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X. et al. (2024) Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* **21**, 1481–1491 https://doi.org/10.1038/s41592-024-02305-7

154 Theodoris, C.V., Xiao, L., Chopra, A., Chaffin, M.D., Al Sayed, Z.R., Hill, M.C. et al. (2023) Transfer learning enables predictions in network biology. *Nature* **618**, 616–624 https://doi.org/10.1038/s41586-023-06139-9 PMID: 37258680

155 Wen, H., Tang, W., Dai, X., Ding, J., Jin, W., Xie, Y, et al. CellPLM: Pre-training of Cell Language Model Beyond Single Cells. *Bioinformatics*. https://doi.org/10.1101/2023.10.03.560734

156 Yang, X., Liu, G., Feng, G., Bu, D., Wang, P., Jiang, J. et al. (2024) GeneCompass: deciphering universal gene regulatory mechanisms with a knowledge-informed cross-species foundation model. *Cell Res.* **34**, 830–845 https://doi.org/10.1038/s41422-024-01034-y

157 Bunne, C., Roohani, Y., Rosen, Y., Gupta, A., Zhang, X., Roed, M. et al. (2024) How to build the virtual cell with artificial intelligence: priorities and opportunities. *Cell* **187**, 7045–7063 https://doi.org/10.1016/j.cell.2024.11.015

158 Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B, et al. (2018) A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309–1324 https://doi.org/10.1016/j.cell.2018.06.052 PMID: 30078704

159 Jagadeesh, K.A., Dey, K.K., Montoro, D.T., Mohan, R., Gazal, S., Engreitz, J.M. et al. (2022) Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* **54**, 1479–1492 https://doi.org/10.1038/s41588-022-01187-9

160 Dann, E., Cujba, A.-M., Oliver, A.J., Meyer, K.B., Teichmann, S.A. and Marioni, J.C. (2023) Precise identification of cell states altered in disease using healthy single-cell references. *Nat. Genet.* **55**, 1998–2008 https://doi.org/10.1038/s41588-023-01523-7

161 Ho, P., Melms, J.C., Rogava, M., Frangieh, C.J., Poźniak, J., Shah, S.B. et al. (2023) The CD58-CD2 axis is co-regulated with PD-L1 via CMTM6 and shapes anti-tumor immunity. *Cancer Cell* **41**, 1207–1221. https://doi.org/10.1016/j.ccell.2023.05.014

162 Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A. et al. (2021) A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 https://doi.org/10.1038/s41588-021-00911-1

163 Maurer, K., Park, C.Y., Mani, S., Borji, M., Raths, F., Gouin, K.H., III et al. (2025) Coordinated immune networks in leukemia bone marrow microenvironments distinguish response to cellular therapy. *Sci. Immunol.* **10**, eadr0782 https://doi.org/10.1126/sciimmunol.adr0782

164 Frangieh, C.J., Melms, J.C., Thakore, P.I., Geiger-Schuller, K.R., Ho, P., Luoma, A.M. et al. (2021) Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nat. Genet.* **53**, 332–341 https://doi.org/10.1038/s41588-021-00779-1 PMID: 33649592

165 Zhao, W., Dovas, A., Spinazzi, E.F., Levitin, H.M., Banu, M.A., Upadhyayula, P. et al. (2021) Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. *Genome Med.* **13**, 82 https://doi.org/10.1186/s13073-021-00894-y PMID: 33975634

166 Dini, A., Barker, H., Piki, E., Sharma, S., Raivola, J., Murumägi, A. et al. (2025) A multiplex single-cell RNA-Seq pharmacotranscriptomics pipeline for drug discovery. *Nat. Chem. Biol.* **21**, 432–442 https://doi.org/10.1038/s41589-024-01761-8

167 Kim, S.Y., De Weert, T.A.E., Vermeulen, M., Ringnalda, F., Kester, L., Zsiros, J, et al. (2025) Organoid drug profiling identifies methotrexate as a therapy for SCCOHT, a rare pediatric cancer. *Sci. Adv* **11**, eadq1724 https://doi.org/10.1126/sciadv.adq1724

168 Kenchappa, R.S., Dovas, A., Argenziano, M.G., Meyer, C.T., Stopfer, L.E., Banu, M.A, et al. (2022) Activation of STAT3 through combined SRC and EGFR signaling drives resistance to a mitotic kinesin inhibitor in glioblastoma. *Cell Rep.* **39**, 110991 https://doi.org/10.1016/j.celrep.2022.110991 PMID: 35732128

169 Calandrini, C. and Drost, J. (2022) Normal and tumor-derived organoids as a drug screening platform for tumor-specific drug vulnerabilities. *STAR Protoc.* **3**, 101079 https://doi.org/10.1016/j.xpro.2021.101079 PMID: 35036959

170 Chramiec, A., Teles, D., Yeager, K., Marturano-Kruik, A., Pak, J., Chen, T. et al. (2020) Integrated human organ-on-a-chip model for predictive studies of anti-tumor drug efficacy and cardiac safety. *Lab Chip* **20**, 4357–4372 https://doi.org/10.1039/d0lc00424c PMID: 32955072

171 Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D.I.S., Zairis, S. et al. (2016) Clonal evolution of glioblastoma under therapy. *Nat. Genet.* **48**, 768–776 https://doi.org/10.1038/ng.3590

172 Torre, E.A., Arai, E., Bayatpour, S., Jiang, C.L., Beck, L.E., Emert, B.L. et al. (2021) Genetic screening for single-cell variability modulators driving therapy resistance. *Nat. Genet.* **53**, 76–85 https://doi.org/10.1038/s41588-020-00749-z PMID: 33398196

173 Harmange, G., Hueros, R.A.R., Schaff, D.L., Emert, B., Saint-Antoine, M., Kim, L.C. et al. (2023) Disrupting cellular memory to overcome drug resistance. *Nat. Commun.* **14**, 7130 https://doi.org/10.1038/s41467-023-41811-8 PMID: 37932277

174 Nadalin, F., Marzi, M.J., Pirra Piscazzi, M., Fuentes-Bravo, P., Procaccia, S., Climent, M. et al. (2024) Multi-omic lineage tracing predicts the transcriptional, epigenetic and genetic determinants of cancer evolution. *Nat. Commun.* **15**, 7609 https://doi.org/10.1038/s41467-024-51424-4

175 Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D. et al. (2019) High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 https://doi.org/10.1038/s41592-019-0548-y PMID: 31501547

176 Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., et al. (2020) High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 https://doi.org/10.1016/j.cell.2020.10.026 PMID: 33188776

177 Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R. et al. (2019) Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 https://doi.org/10.1126/science.aaw1219 PMID: 30923225

178 Abdulraouf, A., Jiang, W., Xu, Z., Zhang, Z., Isakov, S., Raihan, T, et al. (2024) Optics-free Spatial Genomics for Mapping Mouse Brain Aging. *bioRxiv* 2024.08.06.606712 https://doi.org/10.1101/2024.08.06.606712 PMID: 39149282

179 Liao, H., Kottapalli, S., Huang, Y., Chaw, M., Gehring, J., Waltner, O. et al. (2024) Optics-free reconstruction of 2D images via DNA barcode proximity graphs. *bioRxiv* 2024.08.06.606834 https://doi.org/10.1101/2024.08.06.606834 PMID: 39149271

180 Lee, Y., Bogdanoff, D., Wang, Y., Hartoularos, G.C., Woo, J.M., Mowery, C.T. et al. (2021) XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment. *Sci. Adv.* **7** https://doi.org/10.1126/sciadv.abg4755

181 Srivatsan, S.R., Regier, M.C., Barkan, E., Franks, J.M., Packer, J.S., Grosjean, P. et al. (2021) Embryo-scale, single-cell spatial transcriptomics. *Science* **373**, 111–117 https://doi.org/10.1126/science.abb9536

182 Feldman, D., Singh, A., Schmid-Burgk, J.L., Carlson, R.J., Mezger, A., Garrity, A.J, et al. (2019) Optical pooled screens in human cells. *Cell* **179**, 787–799 https://doi.org/10.1016/j.cell.2019.09.016 PMID: 31626775

183 Gu, J., Iyer, A., Wesley, B., Taglialatela, A., Leuzzi, G., Hangai, S. et al. (2025) Mapping multimodal phenotypes to perturbations in cells and tissue with CRISPRmap. *Nat. Biotechnol.* **43**, 1101–1115 https://doi.org/10.1038/s41587-024-02386-x PMID: 39375448

184 Kudo, T., Meireles, A.M., Moncada, R., Chen, Y., Wu, P., Gould, J. et al. (2025) Multiplexed, image-based pooled screens in primary cells and tissues with PerturbView. *Nat. Biotechnol.* **43**, 1091–1100 https://doi.org/10.1038/s41587-024-02391-0 PMID: 39375449