



OPEN ACCESS

EDITED BY
Attila Csikász-Nagy,
Pázmány Péter Catholic University,
Hungary

REVIEWED BY
Dezso Modos,
Quadram Institute, United Kingdom
Andrei Zinovyev,
Institut Curie, France

*CORRESPONDENCE
Bence Szalai,
✉ bence.szalai@turbine.ai

SPECIALTY SECTION
This article was submitted to Translational
Systems Biology and In Silico Trials,
a section of the journal
Frontiers in Systems Biology

RECEIVED 17 December 2022
ACCEPTED 30 January 2023
PUBLISHED 09 February 2023

CITATION
Szalai B and Veres DV (2023), Application
of perturbation gene expression profiles in
drug discovery—From mechanism of
action to quantitative modelling.
Front. Syst. Biol. 3:1126044.
doi: 10.3389/fsysb.2023.1126044

COPYRIGHT
© 2023 Szalai and Veres. This is an open-
access article distributed under the terms
of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Application of perturbation gene expression profiles in drug discovery—From mechanism of action to quantitative modelling

Bence Szalai^{1*} and Dániel V. Veres^{1,2}

¹Turbine Ltd., Budapest, Hungary, ²Department of Molecular Biology, Semmelweis University, Budapest, Hungary

High dimensional characterization of drug targets, compound effects and disease phenotypes are crucial for increased efficiency of drug discovery. High-throughput gene expression measurements are one of the most frequently used data acquisition methods for such a systems level analysis of biological phenotypes. RNA sequencing allows genome wide quantification of transcript abundances, recently even on the level of single cells. However, the correct, mechanistic interpretation of transcriptomic measurements is complicated by the fact that gene expression changes can be both the cause and the consequence of altered phenotype. Perturbation gene expression profiles, where gene expression is measured after a genetic or chemical perturbation, can help to overcome these problems by directly connecting the causal perturbations to their gene expression consequences. In this Review, we discuss the main large scale perturbation gene expression profile datasets, and their application in the drug discovery process, covering mechanisms of action identification, drug repurposing, pathway activity analysis and quantitative modelling.

KEYWORDS

gene expression, perturbation, mechanism of action, drug repurposing, pathway activity, modelling

Introduction

Identification of the systems level alterations in diseases and their relationships to drug effect and efficacy are crucial to better understand drug-disease relationships and develop new therapeutics (X. Yang et al., 2020). The most frequently used high-throughput, genome wide (“omics”) methods for such a systems level characterisation are still transcriptomic measurements such as microarray and RNAseq (McGettigan 2013; Manzoni et al., 2018). Despite the relatively affordable acquisition and well established analysis methods, the correct interpretation of gene expression measurements are complicated by several factors. Classical analysis methods return lengthy lists of differentially expressed genes (e.g., healthy vs. control sample), however differential expression does not necessarily mean altered activity on protein level (Nusinow et al., 2020; Piran et al., 2020), and also differentially expressed genes are frequently not the cause, but the consequence of the investigated phenotype. While different prior-knowledge based bioinformatics methods like pathway analysis techniques (Nguyen et al., 2019) can help in the interpretation, identifying the causal alterations are still difficult.

Perturbation gene expression signatures are defined as the gene expression difference between a perturbed and control condition, calculated by differential expression analysis (Love et al., 2014; Ritchie et al., 2015). In case of perturbation signatures, we can directly connect the cause (perturbation) and the downstream effect (gene expression signature), which can help to

understand cellular mechanisms (Lamb et al., 2006). Perturbation gene expression profiles can be generated on gene level (by knocking-out/down or overexpressing the gene of interest) and also on compound level. Analysis of drug induced perturbation signatures can help in several steps of the drug discovery process (Figure 1). Comparison of drug and gene related gene expression profiles can highlight drug mechanisms of action, and identify potential off-targets. Comparison of disease signatures (differential expression of disease and corresponding healthy tissue) and drug signatures is frequently used to identify new indications for existing drugs. Analysing signatures of drug pairs can identify synergistic and antagonistic drug combinations. Finally, perturbation signatures can help to develop methods to gain mechanistic understanding of cellular processes from omics data.

In this Review, we discuss the most frequently used perturbation transcriptomics measurement methods and the corresponding large scale dataset, and the above mentioned main applications, with a focus on cancer drug discovery.

Perturbation gene expression measurement methods and datasets

While large amounts of low-scale perturbation gene expression datasets are available in various public repositories [like GEO (Barrett et al., 2013) and ArrayExpress (Kolesnikov et al., 2015)], their usability is hindered by the complicated searchability, lack of uniform metadata format and possible batch effects between studies. To overcome these problems, several authors created secondary databases, where gene

expression profiles from these public repositories were collected, uniformly preprocessed and even metadata is standardised (Lachmann et al., 2018). Crowdsourcing methods and natural language processing (NLP) techniques can help to speed up the lengthy manual metadata curation in these cases (Z. Wang et al., 2016). Generally, a large proportion of these collected datasets are focusing on transcriptomics changes in cancer cell lines, but there are also dedicated collections of gene expression signatures from non-cancerous tissues (Zheng et al., 2022).

Connectivity Map (Lamb et al., 2006) was the first large-scale attempt to create a compendium of gene expression signatures. Connectivity Map used originally 164 small molecules as perturbations in 4 cell lines, and gene expression was measured with Affymetrix microarrays. While this dataset was used in several studies investigating drug mechanism of action and repurposing, Connectivity Map lacked both genetic perturbations and diversity of cell lines. High cost of microarray and bulk RNA sequencing made perturbation gene expression profile generation unscalable, thus the authors of the original Connectivity Map created the “next-generation” Connectivity Map by measuring a reduced transcriptome, using the L1000 technology (Subramanian et al., 2017). Hybridisation based L1000 assay measures only the expression of 978 “landmark” genes, and the rest of the transcriptome is computationally inferred. L1000 technology allowed the generation of more than 1,000,000 perturbation gene expression profiles until 2017, and a new >3,000,000 profile dataset is also available (at <https://clue.io/>). Importantly, the new Connectivity Map contains more cell lines and a more diverse collection of perturbations, including genetic (shRNA, CRISPR, and

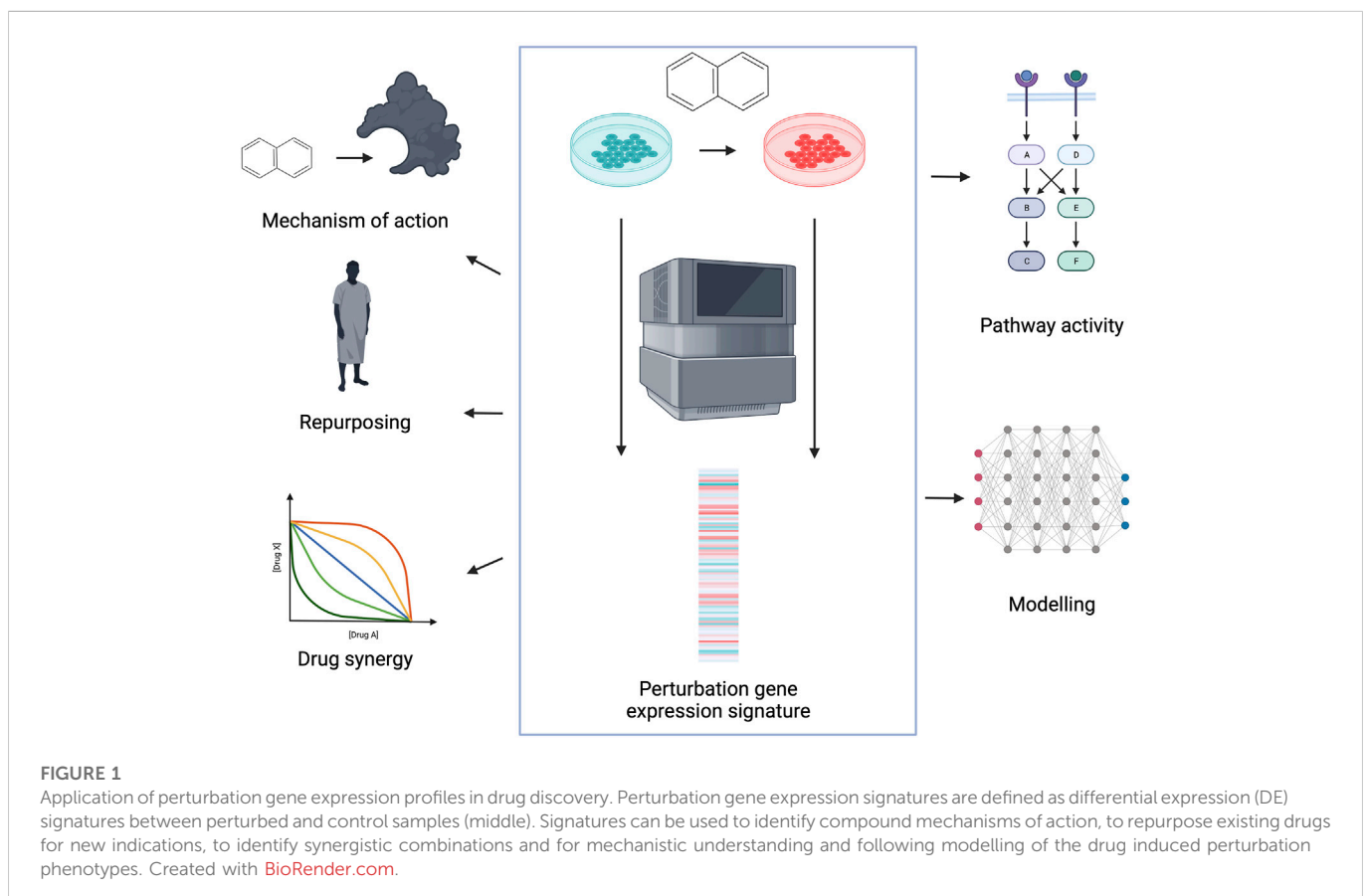


TABLE 1 Public datasets and databases for perturbation gene expression profiles.

Dataset	Perturbation type	Short description	Database URL	Reference
CREEDS	chemical and genetic	crowdsourced collection of perturbation signatures from GEO	https://maayanlab.cloud/CREEDS/	Wang et al. (2016)
ChemPert	chemical	collection of non-cancer perturbation signatures	https://chempert.uni.lu/	Zheng et al. (2022)
Connectivity Map (LINCS)	chemical and genetic	L1000 assay based, >3,000,000 signatures	https://clue.io/	Subramanian et al. (2017)
PANACEA	chemical	anti-cancer drug perturbation signatures in multiple cell lines, RNAseq		Douglass et al. (2022)
Perturb-Seq (genome wide)	genetic	genome-wide CRISPRi in 1 cell line, scRNA-seq	https://gwps.wi.mit.edu/	Replogle et al. (2022)
Mix-Seq	chemical	anticancer drug perturbations in multiple cell lines, scRNA-seq		McFarland et al. (2020)

overexpression), chemical (small molecules) and physiological (ligand) perturbations.

While RNA sequencing was considered too expensive to generate large scale perturbation gene expression profiles, recent technological advancements, especially single cell sequencing methods, brought about a change in this. By using barcoding and pooling strategies, bulk sequencing methods like PLATE-Seq (Bush et al., 2017) and DRUG-Seq (Ye et al., 2018) reached comparable costs to the L1000 assay and allowed the production of large scale drug perturbation screens like PANACEA (Douglass et al., 2022). Recent advancement of single cell RNA sequencing (scRNA-seq) methodologies further increased the throughput of perturbation gene expression profiling. One of the pioneer methods of the field, Perturb-Seq combines CRISPR (Dixit et al., 2016) or CRISPRi (Adamson et al., 2016) and sc-RNAseq with the help of expressed guide barcodes. Perturb-Seq allows genome-wide (Replogle et al., 2022) genetic perturbations, but only in one cell line in each experiment. Other methods, like MIX-Seq (McFarland et al., 2020), use a smaller number of chemical perturbations (anti-cancer drugs), but measure gene expression profiles in a large number of cell lines, with the help of SNP-based computational demultiplexing. Importantly, scRNA-seq based methods can also identify heterogeneity of perturbation response. As the number of scRNA-seq perturbation screens increases, it is important to categorise and harmonise these datasets, as done in (Peidli et al., 2022).

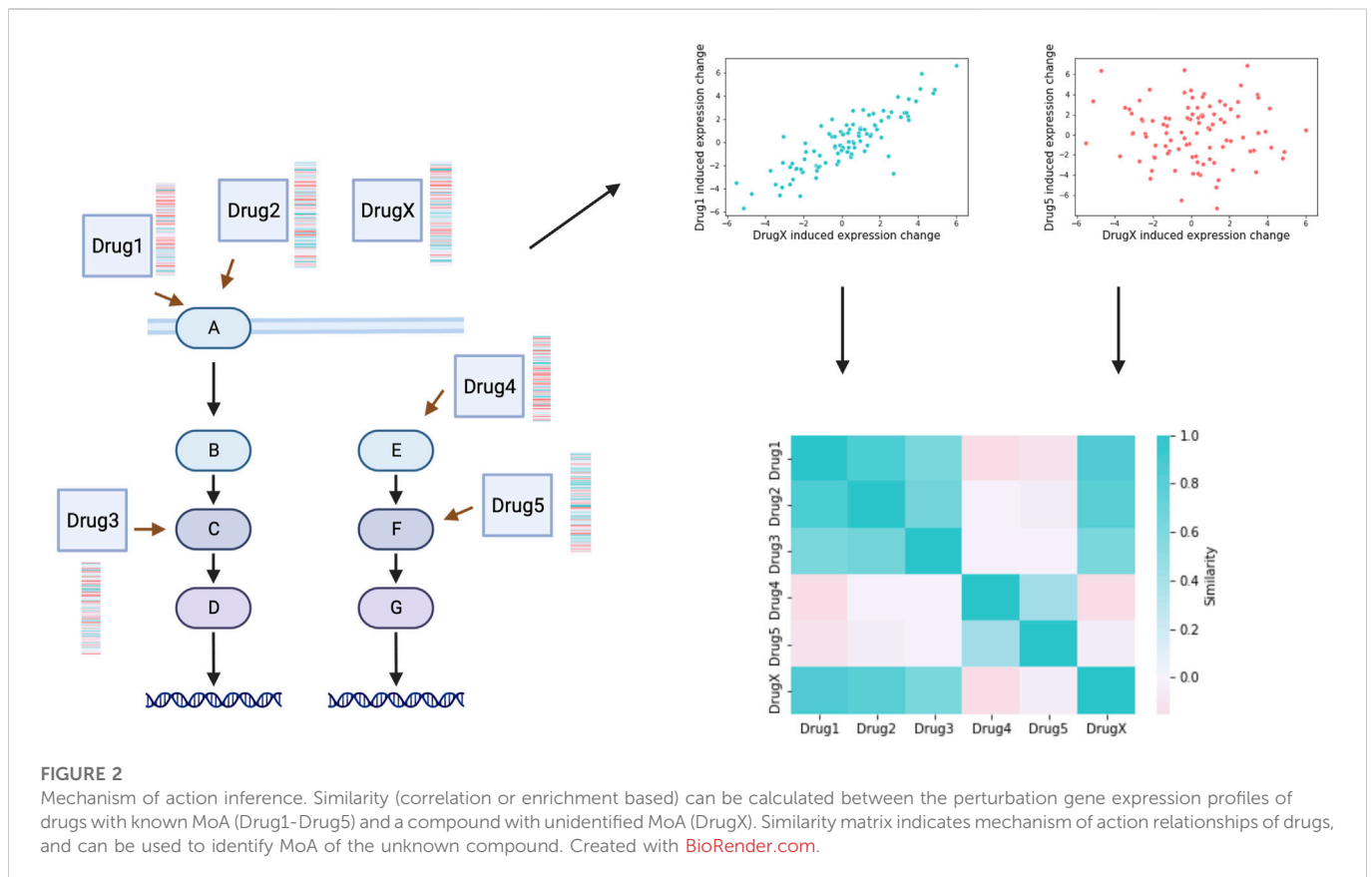
Currently a large number of ultra-high throughput methods are available for new hybridization (L1000) and bulk/single cell RNA-seq based perturbation gene expression screens. Also, the previously described (Table 1), public datasets give rich sources for *in silico* analysis of existing results. However, it is important to highlight that all of these different methods and datasets have their own intrinsic biases, highlighting the importance of harmonisation and comparison of results from different sources. First important factor to consider is the type of the used assay: L1000 measures only a reduced part of the transcriptome, which can only explain partial variance (~90%) of total transcriptomics difference (Subramanian et al., 2017). While the rest of the transcriptome is inferred, the computationally inferred transcriptomic changes are less reliable, especially in case of very specific perturbations (e.g., shRNA induced gene expression decrease of the target gene is generally only detected in case of landmark genes, and not in case of inferred genes). Single cell sequencing also leads to lower number of detected genes than classical bulk sequencing, however pseudo-bulking methods can help to overcome these

problems (McFarland et al., 2020). While lower coverage in case of L1000 and scRNAseq based assays can be problematic to identify specific differentially expressed genes, multigene signature based techniques (discussed in the following sections) are less sensitive for the lower number of detected genes (Holland et al., 2020). It is also important to consider perturbation type related differences in case of interpretation. While CRISPR has the highest specificity, shRNA perturbations can lead to partial inhibition, which resembles drug effect better in some cases (Michael Krill-Burger et al., 2022). In case of drug perturbations, increasing the drug concentration can lead to higher proportion of off-target effects, and (especially in case of oncology drugs) increased toxicity, which can mask the compound specific transcriptional effect (Szalai et al., 2019).

Mechanism of action inference

As potential off-targets can influence both adverse effects and clinical efficacy (Lin et al., 2019), characterisation of drug target profiles is a crucial step of drug development. Also, identifying new targets of existing drugs can facilitate drug repurposing in new indications (see also next section). Classical methods characterise drug targets by the binding strength of drugs to individual target proteins. While these methods can effectively characterise the binding characteristics to the main targets (or shortlisted off-target candidates), they are not feasible on genome/proteome scale to identify off-targets. In contrast, gene expression changes induced by drug perturbations can help to define target profiles potentially on the genome scale (Trapotsi et al., 2022).

The basic principle of gene expression profile based mechanism of action (MoA) identification is based on the fact that compounds with shared mechanisms of action lead to similar changes of cellular signalling mechanisms, thus leading to similar gene expression changes (Figure 2). The perturbation gene expression profile of a drug with unknown mechanism of action can be used to query large scale datasets (see previous section) and identify potential MoA based on similarities (Musa et al., 2018). Frequently used similarity metrics are generally correlation (Szalai et al., 2019) or enrichment (Subramanian et al., 2017) based. Drug signature based MoA identifications has successfully used to identify Rho-kinase inhibitor *Fasudil* as autophagy inducer (Iorio et al., 2010), PKC inhibitor *Enzastaurin* as GSK3 inhibitor (Subramanian et al., 2017) and the



role of JAK2 in the MoA of *Mitomycin C* (Woo et al., 2015). In a recent large scale, crowdsourced benchmarking study (Douglass et al., 2022), participants used perturbation gene expression signatures of cancer cell lines treated with 32 kinase inhibitors to predict the targets of these (for the participants unknown) drugs. Best performing methods were able to predict experimentally verified targets with ROC AUC >0.7, also confirming the applicability of gene expression signatures for genome wide inference of drug (off-) targets. With this study (Douglass et al., 2022), also created a benchmark dataset for further computational studies.

Drug signature based MoA identification has similar performance as the gold standard of the field, chemical similarity based methods (Baillif et al., 2020). However, while chemical similarity based target/MoA prediction can be performed on any compound of a (even virtual) library, expression profile similarity based predictions require prior measurement of drug induced gene expression signature. On the other hand, expression signatures can help to identify more “unexpected” off-targets and mechanism of actions. This is especially true, if the similarity is calculated between the signatures of chemical and genetic perturbations, where MoA is identified not through “guilt-by-association” (similarity to drugs with known MoA), but based on the direct similarity of a drug’s and its target’s genetic perturbation profile. Recently several methods were developed to infer compound MoA based on CRISPR and shRNA induced signatures (Pabon et al., 2019; Jang et al., 2021; Zhong et al., 2022). On the other side, time scale and efficacy of perturbation can be substantially different between chemical and genetic perturbations (e.g., in Connectivity Map (LINCS) dataset

gene expression is measured 24 h after drug, but 96 h after genetic perturbation), which can make cross perturbation modality comparisons more complicated. Importantly, perturbation signatures can be generated for each investigated cell line, but “consensus” (i.e., general, cell line independent) signatures of drugs can be also created. While using consensus signatures can simplify analysis pipelines, they can mask cell line specific effects (Innes and Bader 2021), and as recent analysis suggest, some methods for consensus signature calculation can lead to artificial similarities of unrelated signatures (Smith and Haibe-Kains 2022). Also, it is important to highlight that similarity between gene expression profiles does not necessarily imply shared MoA, especially in case of anti-cancer drugs. Anti-cancer drugs lead to decreased cell viability, which is represented in their perturbation gene expression signatures (Szalai et al., 2019; Jones et al., 2020; McFarland et al., 2020). Thus two cytotoxic drugs can have similar gene expression signature, despite having distinct mechanisms of action, which effect can be removed by appropriate statistical models (Szalai et al., 2019; McFarland et al., 2020). On the other side, drug and genetic perturbation induced cell viability changes can be also used to identify the target profile of anti-cancer compounds: correlating drug sensitivity (Iorio et al., 2016; Corsello et al., 2020) and gene essentiality (Tsherniak et al., 2017; Behan et al., 2019) on large panels of cancer cell lines (Ghandi et al., 2019; van der Meer et al., 2019) can help to identify on- and off-targets based on sensitivity profile similarity (W. Wang et al., 2022; Gonçalves et al., 2020). Gene expression signature and sensitivity profile based methods can complement each other in computational MoA inference.

Drug repurposing—Signature reversion

Drug repurposing, the process of finding new indications for existing, approved drugs, gains more and more relevance with the increasing costs of *de novo* drug development (Pushpakom et al., 2019). Importantly, already approved drugs have a lower chance to fail due to toxicity in new indications. Drug repurposing is especially important in case of rare diseases, where small market size makes *de novo* drug discovery even more complicated.

Drug induced perturbation gene expression profiles are frequently used for computational drug repurposing (Sirota et al., 2011). The main hypothesis behind signature reversal based methods is, that if a drug induced gene expression signature is anti-similar to a disease related gene expression signature, then the drug can potentially reverse the disease specific gene expression changes, thus the disease phenotype. In these studies, the similarity of disease signatures and drug induced signatures is calculated, and drugs showing negative similarity are prioritised for further experimental validation. Signature reversal hypothesis - despite its relative simplicity—led to identification of repurposable drug candidates from cancer (B. Chen et al., 2017; Stathias et al., 2018) through inflammatory (Malcomson et al., 2016) to metabolic diseases (Kunkel et al., 2011).

While signature reversion is frequently used to identify anti-cancer compounds (B. Chen et al., 2017), recently the confounding role of cell proliferation in these studies have been revealed. While cancer signature (differential expression signature between cancer samples and corresponding healthy cells/tissues) contains a strong cell proliferation related component, anti-cancer drugs generally inhibit cell proliferation, and their gene expression signature contains a strong anti-proliferative (cell death related) component (Szalai et al., 2019). This suggests that the anti-similarity of cancer and anti-proliferative drug signatures is trivial, and the drugs identified by signature reversal methods are not necessarily effective in the investigated cancer type, just general toxic compounds. A recent publication (Koudijs et al., 2022) showed that removing the confounding effect of proliferation/anti-proliferation related gene expression changes significantly decreased the predictive performance of signature reversal methods.

Another disease indication, where the signature reversal hypothesis has to be used with caution, is infectious diseases. During the COVID-19 pandemic, signature reversal methods have been frequently used to identify potential antiviral drugs against SARS-CoV-2. Several of these studies found that drugs having a similar (and not anti-similar, which would be assumed based on the signature reversal hypothesis) gene expression signature to SARS-CoV-2 infection induced transcriptomics signature are effective *in vitro* antivirals (F. Chen et al., 2021; Barsi et al., 2022). As in case of viral infection diseases, infected host cells activate adaptive, antiviral pathways (like NFkB and JAK-STAT), supporting and not reversing these activities indeed can have beneficial, antiviral effects. Nevertheless, other studies found that drugs showing anti-similarity to influenza infection signature are effective antivirals (Pizzorno et al., 2019), thus the general usefulness of signature reversal methods in infection diseases needs further evaluation.

Most current signature reversal methods use bulk disease transcriptomics data for drug prioritisation. However, in a bulk tissue sample, several different cell types exist. While the gene expression changes of some of these cells can have a causal role in

disease development (thus are candidates for signature reversing drugs), other cell types' gene expression profile can change as a consequence of disease process. Using single cell RNA-seq to identify cell type specific disease signatures and repurposable drugs can further increase the applicability of signature reversal methods (Liu et al., 2022).

Identifying synergistic drug combinations

Using drug combinations can help to use lower drug doses, thus can decrease the frequency of adverse effects, and can help to overcome drug resistance mechanisms, especially in case of anti-cancer compounds. Drug combinations are classified as synergistic, additive or antagonistic, based on the difference between observed and expected drug effects, where the expected drug effect is calculated using some synergy model like Bliss independence or Loewe additivity models (Yadav et al., 2015). To experimentally measure synergy, multiple dose—Response curve measurements are required, thus large-scale experimental testing is generally not feasible due to the combinatorial increase of the number of possible combinations. Computational methods are frequently used to infer drug synergy for new combinations in new biological samples.

Generally, machine learning models (Menden et al., 2019) use features of drugs, and features of the cell line to predict synergistic effects of the drugs. The most frequently used drug features are chemical fingerprints or other representation of the drugs chemical structure (Preuer et al., 2018). However, drug induced gene expression signatures contain more context specific biological information regarding drug effect than chemical structure, thus their application for drug synergy prediction can be beneficial. Using drug signatures has been used to predict synergistic effects of anti-cancer drugs (Bansal et al., 2014), and a recent benchmarking study showed that machine learning models using expression based features significantly outperform standard, chemical feature based methods (El Khili et al., 2022). Interestingly, several of these studies suggest that similarity between drug signatures is a strong predictor of synergistic drug effect (Diaz et al., 2020), suggesting drugs targeting the same pathway, but at different targets are generally more synergistic. Other studies found that strong compound similarities, but also dissimilarities are correlating with synergistic drug effect (M. Yang et al., 2020), suggesting that more detailed computational studies and experimental datasets are needed to fully understand mechanisms behind drug synergy. Importantly, measuring combination induced transcriptomics changes (Mathur et al., 2022) can help to resolve these ambiguities.

Interpretation—Pathway activity

While mechanism of action, drug repurposing, and even synergy prediction is possible by similarity based/black-box machine learning methods, biological, mechanistic interpretation of drug induced gene expression changes can also help to better understand drug effect and identify potential biomarkers. Standard methods for such a biological interpretation are gene set enrichment and pathway analysis techniques (Nguyen et al., 2019).

Classical pathway analysis techniques calculate some kind of enrichment of differentially expressed genes, using pathway

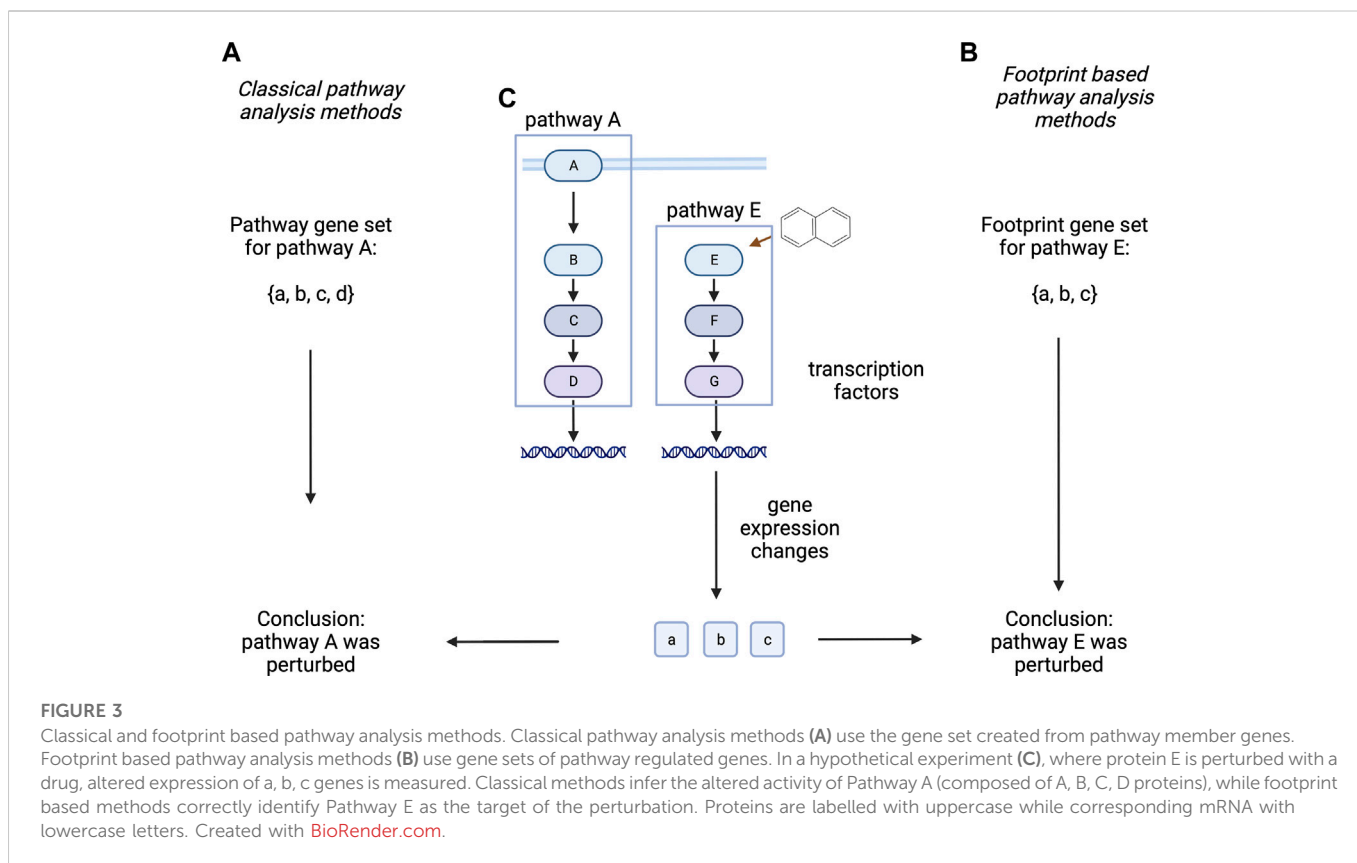


TABLE 2 Computational tools for footprint based pathway activity inference.

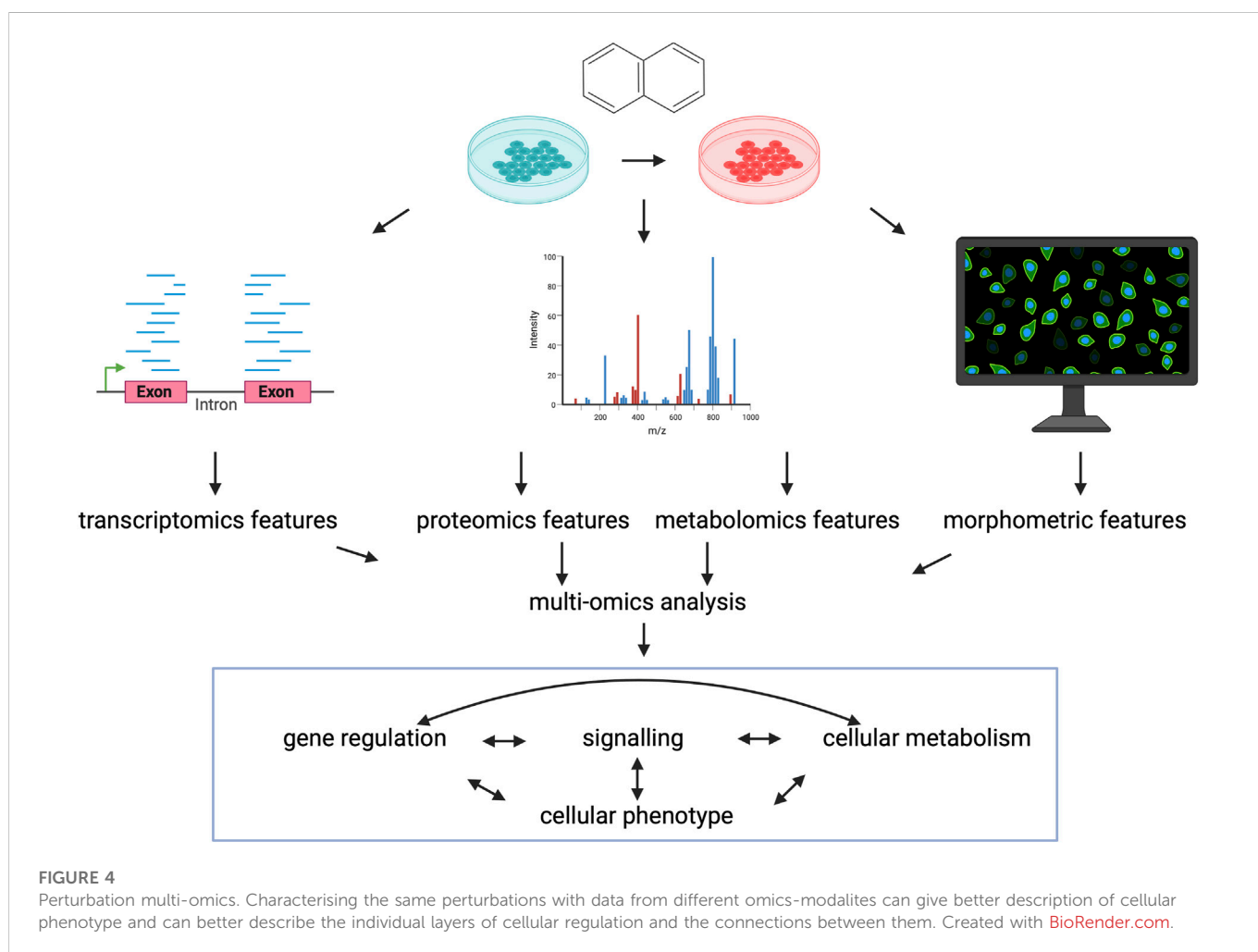
Tool	Inferred activity	Short description	Tool URL	Reference
ChEA3	Transcription factor	Co-expression and ChIP-seq based regulatory interactions for TFs	https://maayanlab.cloud/chea3/	Keenan et al. (2019)
CytoSig	Cytokine	Signatures collected for 43 cytokines form perturbation data	https://cytosig.ccr.cancer.gov/	Jiang et al. (2021)
DoRothEA	Transcription factor	Co-expression, motif, ChIP-seq and literature based regulatory interactions for TFs	https://bioconductor.org/packages/release/data/experiment/html/dorothea.html	Garcia-Alonso et al. (2019)
Viper	Protein	Co-expression based regulatory interactions	https://www.bioconductor.org/packages/release/bioc/html/viper.html	Alvarez et al. (2016)
PROGENy	Pathway	Signatures collected for 14 cancer related pathways from perturbation data	https://bioconductor.org/packages/release/bioc/html/progeny.html	Schubert et al. (2018); Holland et al. (2019)
SPEED2	Pathway	Signature collected for 16 pathways from perturbation data	https://speed2.sys-bio.net/	Rydenfelt et al. (2020)

member genes as gene sets (Subramanian et al., 2005; Liberzon et al., 2015). Importantly, these methods (indirectly) assume some clear connection between gene expression, protein abundance and protein activity, however these assumptions are not necessarily correct (Szalai and Saez-Rodriguez 2020). Other methods are using the expression of pathway regulated (“footprint”) genes to infer which pathway activity changes had led to the observed gene expression pattern (A. Dugourd and Saez-Rodriguez 2019). These later methods have been shown to better represent biological phenotype in several benchmarks than classical,

pathway membership based pathway analysis techniques (Cantini et al., 2018; Holland et al., 2019; Szalai and Saez-Rodriguez 2020; Douglass et al., 2022). One can argue that “footprint” based methods are better suited to identify pathway activity changes responsible for the gene expression changes of a sample, while classical, membership based methods try to infer the possible consequence of gene expression changes on pathway activity. While the latter can identify important information (like negative feedback mechanisms), it is generally more speculative than results of “footprint” based methods (Figure 3).

TABLE 3 Gene expression based mechanistic models.

Tool	Short description	Tool URL	Reference
CARNIVAL	Identifies signalling pathway activity changes by connecting perturbations to inferred transcription factor activities using integer linear programming.	https://github.com/saezlab/CARNIVAL	Liu et al. (2019)
NicheNet	Predicts protein (ligand-receptor) activity by connecting receptors to gene expression through signalling and gene regulatory network. Strength of ligand—gene expression interactions are calculated <i>via</i> Personalised PageRank.	https://github.com/saeyslab/nichenetr	Browaeys et al. (2019)
KPNN	Predicts protein activity from gene expression using a neural network architect resembling gene regulatory and signalling network.	https://github.com/epigen/KPNN	Fortelny and Bock. (2020)
CausalPath	Identifies causal priors (causal graph motifs) from Pathway Commons database Cerami et al. (2011), and matches them with correlated changes from the analysed data. Uses also (phospho)proteomics data	https://github.com/PathwayAndDataAnalysis/causalpath	Babur et al. (2021)
LEMBAS	Predicts transcription factor activity from ligand stimulation using a recurrent neural network resembling signalling network.	https://github.com/Lauffenburger-Lab/LEMBAS	Nilsson et al. (2022)



Of course, to apply “footprint” based pathway analysis methods, prior knowledge regarding the pathway regulated genes is required. This information can be collected *via* literature mining and gene regulatory network inference (Alvarez et al., 2016; Garcia-Alonso et al., 2019; Keenan et al., 2019), but perturbation expression profile datasets are also excellent sources for this information.

Methods like PROGENy (Schubert et al., 2018) or SPEED (Parikh et al., 2010; Rydenfelt et al., 2020) collected a large set of gene expression profiles related to the investigated pathways, and used different statistical models to identify the pathway responsible “footprint” genes (Table 2). Footprint concept has recently extended to infer ligand/receptor associations and activities (Jiang

et al., 2021). Interestingly, benchmarking studies suggest (Badia-i-Mompel et al., 2022) that quality of the used footprint gene set has a higher influence on the performance, than the used statistical method. In case of transcription factor activity inference methods, the ones using multiple sources of regulatory interactions (like ChIP-Seq data, co-expression, literature curated data and promoter binding motifs) perform better (Garcia-Alonso et al., 2019; Keenan et al., 2019) than the ones relying on single sources of information. The performance of perturbation profile based pathway activity inference methods is also strongly dependent on the quality and amount of the collected perturbation experiments (Schubert et al., 2018). Perturbation expression profiles are also needed for the correct benchmarking of pathway analysis techniques, as for assessing a newly developed method, ground-truth data is required. Importantly, while for the development and benchmarking of footprint based pathway activity inference tools perturbation gene expression data is suitable, these methods can be effectively used also on baseline (e.g., disease samples) expression data (Schubert et al., 2018).

Modelling cellular phenotype

One of the main goals of systems biology studies is to develop mechanistically understandable and simulatable models of cellular processes (Froehlich et al., 2017). Importantly, virtual perturbation in these models can help to identify biomarkers and synergistic drug combinations (Eduati et al., 2017). These models generally use some prior-knowledge biological (signalling) network (Türei et al., 2021) to connect proteins (nodes) and use data-driven methods to parameterise (fit) the network parameters (edge weights) to the biological context (Garrido-Rodriguez et al., 2022). To fit the network parameters, a wide range of computational tools are used, like graph algorithms (Browaeys et al., 2019), integer linear programming (Liu et al., 2019) or neural network architectures (Yuan et al., 2021; Nilsson et al., 2022). Perturbation data is especially suitable for contextualisation of these models (Korkut et al., 2015), as in case of perturbation experiments the response of the same cellular system is measured, which lowers the amount of possible parameters of the model.

While proteomics data has been used most frequently for simulation of signalling networks (Yuan et al., 2021), recently gene expression data has been also effectively applied in this context (Table 3). Importantly, while (phospho) proteomics data can be used directly to approximate activity of signalling components, transcriptomics data is less suitable for modelling signal flow. However, either by using gene expression data in gene regulatory context (Browaeys et al., 2019; Liu et al., 2019; Fortelny and Bock 2020; Babur et al., 2021) or inferring protein activities from gene expression (Nilsson et al., 2022) can help to build effective mechanistic models for cell signalling using transcriptomics data. While perturbation data based mechanistic models are obviously hard to create for patient data, baseline data could also be used to transfer and contextualise these models to *in vivo* settings (Saez-Rodriguez and Blüthgen 2020). Currently, the main application of these modelling frameworks is hypothesis generation and more general and unbiased benchmarks are needed to compare them and measure their general predictive performance.

Conclusion

As collected above, perturbation gene expression profiles can give rich input data for drug discovery and development. Drug induced expression changes can help to identify the on- and off-targets of newly developed compounds, comparison to disease signatures can reveal new disease indications, and suitable analysis of gene expression signatures can give mechanistic insight regarding drug action.

Additionally, recently more and more perturbation data from other omics modalities is available, suggesting the importance of perturbation multi-omics in the future. High throughput, high content imaging and featurization of images allows to derive morphological signatures of perturbed cell states, which can be used to identify compound mechanisms of action similarly to gene expression profiles (Haghighi et al., 2022). Also, morphological profiles are generally interpretable for cell biologists, and can be directly connected to perturbation induced cell states. Proteomics and phospho-proteomics are more closely connected to the activity of cellular process than gene expression, so perturbation proteomics datasets can be used to infer compound effects of signalling and compound similarities (Zhao et al., 2020; Gabor et al., 2021). Recently also drug induced metabolomics changes were used to describe cellular phenotype and synergistic drug effect (Lu et al., 2022). Figure 4.

Most importantly, different omics layers can measure different, not interchangeable variance of cellular phenotype (Gross et al., 2022), thus using integrative data analysis (Argelaguet et al., 2020) and modelling methods (Dugourd et al., 2021) of different, but harmonised omics modalities can lead to better understating of drugs' effect on cellular processes in the future.

In summary, perturbation gene expression measurements create valuable insight to analyse drug effect in target cells, and the computational tools developed with perturbation data can also be generalised to baseline, including patient data.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors thank Miklós Laczik for the valuable feedback on the manuscript.

Conflict of interest

DV is founder, BS and DV are employees of Turbine Ltd.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- 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 (7), 1867–1882. doi:10.1016/j.cell.2016.11.048
- Alvarez, M. J., Shen, Y., Giorgi, F. M., Ding, B. B., Ye, B. H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 48 (8), 838–847. doi:10.1038/ng.3593
- Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., et al. (2020). MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21 (1), 111. doi:10.1186/s13059-020-02015-1
- Babur, Ö., Korkut, A., Durupinar, F., Siper, C., Dogrusoz, U., Vaca Jacome, A. S., et al. (2021). Causal interactions from proteomic profiles: Molecular data meet pathway knowledge. *Patterns (New York, N.Y.)* 2 (6), 100257. doi:10.1016/j.patter.2021.100257
- Badia-i-Mompel, P., Vélez Santiago, J., Jana, B., Geiss, C., Daniel, D., Müller-Dott, S., et al. (2022). decoupleR: Ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* 2 (1), vbac016. doi:10.1093/bioadv/vbac016
- Baillif, B., Wichard, J., Méndez-Lucio, O., and Rouquié, D. (2020). Exploring the use of compound-induced transcriptomic data generated from cell lines to predict compound activity toward molecular targets. *Front. Chem.* 8, 296. doi:10.3389/fchem.2020.00296
- Bansal, M., Yang, J., Karan, C., Menden, M. P., Costello, J. C., Tang, H., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* 32 (12), 1213–1222. doi:10.1038/nbt.3052
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for functional genomics data sets--update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193
- Barsi, S., Papp, H., Valdeolivas, A., Tóth, D. J., Kuczmog, A., Madai, M., et al. (2022). Computational drug repurposing against SARS-CoV-2 reveals plasma membrane cholesterol depletion as key factor of antiviral drug activity. *PLoS Comput. Biol.* 18 (4), e1010021. doi:10.1371/journal.pcbi.1010021
- Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR–cas9 screens. *Nature* 568, 511–516. doi:10.1038/s41586-019-1103-9
- Browaers, R., Saelens, W., and Saeys, Y. (2019). NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat. Methods*, 17, 159–162. doi:10.1038/s41592-019-0667-5
- 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 (1), 105. doi:10.1038/s41467-017-00136-z
- Cantini, L., Calzone, L., Martignetti, L., Rydenfelt, M., Blüthgen, N., Barillot, E., et al. (2018). Classification of gene signatures for their information value and functional redundancy. *NPJ Syst. Biol. Appl.* 4, 2. doi:10.1038/s41540-017-0038-8
- Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., et al. (2011). Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, D685–D690. doi:10.1093/nar/gkq1039
- Chen, B., Ma, L., Paik, H., Sirota, M., Wei, W., Chua, M. S., et al. (2017). Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat. Commun.* 8, 16022. doi:10.1038/ncomms16022
- Chen, F., Shi, Q., Pei, F., Vogt, A., Porritt, R. A., Garcia, G., et al. (2021). A systems-level study reveals host-targeted repurposable drugs against SARS-CoV-2 infection. *Mol. Syst. Biol.* 17 (8), e10239. doi:10.15252/msb.202110239
- Corsello, S. M., Nagari, R. T., Spangler, R. D., Jordan, R., Kocak, M., Bryan, J. G., et al. (2020). Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* 1 (2), 235–248. doi:10.1038/s43018-019-0018-6
- Diaz, J. E., Eren Ahsen, M., Schaffter, T., Chen, X., Realubit, R. B., Karan, C., et al. (2020). The transcriptomic response of cells to a drug combination is more than the sum of the responses to the monotherapies. *eLife* 9, e52707. doi:10.7554/eLife.52707
- 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 (7), 1853–1866. doi:10.1016/j.cell.2016.11.038
- Douglass, E. F., Jr, Allaway, R. J., Szalai, B., Wang, W., Tian, T., Fernandez-Torras, A., et al. (2022). Community challenge for a pancreatic drug mechanism of action inference from perturbational profile data. *Cell. Rep. Med.* 3 (1), 100492. doi:10.1016/j.xcrm.2021.100492
- Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Gabor, A., Emdal, K. B., et al. (2021). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Mol. Syst. Biol.* 17 (1), e9730. doi:10.15252/msb.20209730
- Dugourd, A., and Saez-Rodriguez, J. (2019). Footprint-based functional analysis of multi-omic data. Available At: <https://www.sciencedirect.com/science/article/pii/S2452310019300149>.
- Eduati, F., Doldán-Martelli, V., Klinger, B., Cokelaer, T., Sieber, A., Kogera, F., et al. (2017). Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. *Cancer Res.* 77 (12), 3364–3375. doi:10.1158/0008-5472.CAN-17-0078
- El Khili, M. R., Memon, S. A., and Amin, E. (2022). Marsy: A multitask deep learning framework for prediction of drug combination synergy scores. bioRxiv. doi:10.1101/2022.06.07.495155
- Fortelny, N., and Bock, C. (2020). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biol.* 21 (1), 190. doi:10.1186/s13059-020-02100-5
- Froehlich, F., Kessler, T., Weindl, D., Shadrin, A., Leonard, S., Hache, H., et al. (2017). Efficient parameterization of large-scale mechanistic models enables drug response prediction for cancer cell lines. bioRxiv. doi:10.1101/174094
- Gabor, A., Tognetti, M., Driessen, A., Tanevski, J., Guo, B., Cao, W., et al. (2021). Cell-to-Cell and type-to-type heterogeneity of signaling networks: Insights from the crowd. *Mol. Syst. Biol.* 17 (10), e10402. doi:10.15252/msb.202110402
- García-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., Saez-Rodriguez, J., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 29 (8), 1363–1375. doi:10.1101/gr.240663.118
- Garrido-Rodriguez, M., Zirngibl, K., Ivanova, O., Lobentzner, S., Saez-Rodriguez, J., and Saez-Rodriguez, J. (2022). Integrating knowledge and omics to decipher mechanisms via large-scale models of signaling networks. *Mol. Syst. Biol.* 18 (7), e11036. doi:10.15252/msb.202211036
- Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Christopher, C., McDonald, E. R., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature* 569 (7757), 503–508. doi:10.1038/s41586-019-1186-3
- Gonçalves, E., Segura-Cabrera, A., Clare, P., Picco, G., Behan, F. M., Jaaks, P., et al. (2020). Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens. *Mol. Syst. Biol.* 16 (7), e9405. doi:10.15252/msb.20199405
- Gross, S. M., Dane, M. A., Smith, R. L., Devlin, K. L., McLean, I. C., Derrick, D. S., et al. (2022). A multi-omic analysis of MCF10A cells provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses. *Commun. Biol.* 5 (1), 1066. doi:10.1038/s42003-022-03975-9
- Haghighi, Ma, Caicedo, J. C., Cimini, B. A., Carpenter, A. E., and Singh, S. (2022). High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nat. Methods* 19 (12), 1550–1557. doi:10.1038/s41592-022-01667-0
- Holland, C. H., Szalai, B., and Saez-Rodriguez, J. (2019). Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochimica Biophysica Acta, Gene Regul. Mech.* 1863, 194431. doi:10.1016/j.bbaggm.2019.194431
- Holland, C. H., Gleixner, J., Kumar, M. P., Mereu, E., and Joughin, B. A. (2020). Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* 21 (1), 36. doi:10.1186/s13059-020-1949-z
- Innes, B. T., and Bader, G. D. (2021). Transcriptional signatures of cell-cell interactions are dependent on cellular context. bioRxiv. doi:10.1101/2021.09.06.459134
- Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., et al. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U. S. A.* 107 (33), 14621–14626. doi:10.1073/pnas.1000138107
- Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell*. 166 (3), 740–754. doi:10.1016/j.cell.2016.06.017
- Jang, G., Park, S., Lee, S., Kim, S., Park, S., and Kang, J. (2021). Predicting mechanism of action of novel compounds using compound structure and transcriptomic signature coembedding. *Bioinformatics* 37 (1), i376–i382. doi:10.1093/bioinformatics/btab275
- Jiang, P., Zhang, Y., Ru, B., Yuan, Y., Vu, T., Paul, R., et al. (2021). Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat. Methods* 18 (10), 1181–1191. doi:10.1038/s41592-021-01274-5
- Jones, A., Tsherniak, A., and McFarland, J. M. (2020). Post-perturbational transcriptional signatures of cancer cell line vulnerabilities. bioRxiv. doi:10.1101/2020.03.04.976217
- Keenan, A. B., Torre, D., Alexander, L., Leong, A. K., Wojciechowicz, M. L., Utti, V., et al. (2019). ChEA3: Transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* 47 (W1), W212–W224. doi:10.1093/nar/gkz446
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Amy Tang, Y., Williams, E., et al. (2015). ArrayExpress update--simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116. doi:10.1093/nar/gku1057
- Korkut, A., Wang, W., Demir, E., Jing, X., Evan, J., Babur, Ö., et al. (2015). Perturbation biology nominates upstream-downstream drug combinations in RAF inhibitor resistant melanoma cells. *eLife* 4, e04640. doi:10.7554/eLife.04640

- Koudijs, K. K. M., Böhringer, S., and Guchelaar, H. K. (2022). Validation of transcriptome signature reversion for drug repurposing in oncology. *Briefings Bioinforma.* 24, bbac490. doi:10.1093/bib/bbac490
- Kunkel, S. D., Suneja, M., Ebert, M., Fox, D. K., Malmberg, S. E., Shields, R. K., et al. (2011). mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. *Cell. Metab.* 13 (6), 627–638. doi:10.1016/j.cmet.2011.03.020
- Lachmann, A., Torre, D., Keenan, A. B., Jagodnik, K. M., Lee, H. J., Wang, L., et al. (2018). Massive mining of publicly available RNA-seq data from human and mouse. *Nat. Commun.* 9 (1), 1366. doi:10.1038/s41467-018-03751-6
- 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 (5795), 1929–1935. doi:10.1126/science.1132939
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The molecular signatures database (MSigDB) hallmark gene set collection. *Cell. Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004
- Lin, A., Giuliano, C. J., Palladino, A., John, K. M., Abramowicz, C., Lou Yuan, M., et al. (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical Trials. *Sci. Transl. Med.* 11 (509), eaaw8412. doi:10.1126/scitranslmed.aaw8412
- Liu, A., Lee, J. H., and Han, N. (2022). scRNA-seq-based drug repurposing targeting idiopathic pulmonary fibrosis (IPF). bioRxiv. doi:10.1101/2022.09.17.508360
- Liu, A., Trairatphisan, P., Gjerga, E., Didangelos, A., Barratt, J., and Saez-Rodriguez, J. (2019). From expression footprints to causal pathways: Contextualizing large signaling networks with CARNIVAL. *NPJ Syst. Biol. Appl.* 5, 40. doi:10.1038/s41540-019-0118-z
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Lu, X., Hackman, G. L., Saha, A., Singh Rathore, A., Collins, M., Friedman, C., et al. (2022). Metabolomics-based phenotypic screens for evaluation of drug synergy via direct-infusion mass spectrometry. *iScience* 25 (5), 104221. doi:10.1016/j.isci.2022.104221
- Malcomson, B., Wilson, H., Veglia, E., Thillaiyampalam, G., Ruth, B., Donegan, S., et al. (2016). Connectivity mapping (csCMap) to predict A20-inducing drugs and their antiinflammatory action in cystic fibrosis. *Proc. Natl. Acad. Sci. U. S. A.* 113 (26), E3725–E3734. doi:10.1073/pnas.1520289113
- Manzoni, C., Kia, D. A., Jana, V., Hardy, J., Wood, N. W., Lewis, P. A., et al. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings Bioinforma.* 19 (2), 286–302. doi:10.1093/bib/bbw114
- Mathur, L., Szalai, B., Du, N. H., Utharala, R., Ballinger, M., Landry, J. J. M., et al. (2022). Combi-seq for multiplexed transcriptome-based profiling of drug combinations using deterministic barcoding in single-cell droplets. *Nat. Commun.* 13 (1), 4450. doi:10.1038/s41467-022-32197-0
- McFarland, J. M., Paoletta, 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 (1), 4296–4315. doi:10.1038/s41467-020-17440-w
- McGettigan, P. A. (2013). Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* 17 (1), 4–11. doi:10.1016/j.cbpa.2012.12.008
- Menden, M. P., Wang, D., Mason, M. J., Szalai, B., Bulusu, K. C., Guan, Y., et al. (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat. Commun.* 10 (1), 2674. doi:10.1038/s41467-019-09799-2
- Michael Krill-Burger, J., Dempster, J. M., Borah, A. A., Paoletta, B. R., Root, D. E., Golub, T. R., et al. (2022). Partial gene suppression improves identification of cancer vulnerabilities when CRISPR-cas9 knockout is pan-lethal. bioRxiv. doi:10.1101/2022.03.02.482624
- Musa, A., Zhang, S. D., Glazko, G., Yli-Harja, O., Dehmer, M., and Haibe-Kains, B. (2018). A Review of connectivity Map and computational approaches in pharmacogenomics. *Briefings Bioinforma.* 19 (3), 506–523. doi:10.1093/bib/bbw112
- Nguyen, T. M., Shafi, A., Nguyen, T., and Draghici, S. (2019). Identifying significantly impacted pathways: A comprehensive Review and assessment. *Genome Biol.* 20 (1), 203. doi:10.1186/s13059-019-1790-4
- Nilsson, A., Peters, J. M., Meimetis, N., Bryson, B., and Lauffenburger, D. A. (2022). Artificial neural networks enable genome-scale simulations of intracellular signaling. *Nat. Commun.* 13 (1), 3069. doi:10.1038/s41467-022-30684-y
- Nusinow, D. P., Szpyt, J., Ghandi, M., Rose, C. M., McDonald, E. R., 3rd, Kalocsay, M., et al. (2020). Quantitative proteomics of the cancer cell line encyclopedia. *Cell.* 180 (2), 387–402. doi:10.1016/j.cell.2019.12.023
- Pabon, N. A., Zhang, Q., Cruz, J. A., Schipper, D. L., Camacho, C. J., and Lee, R. E. C. (2019). A network-centric approach to drugging TNF-induced NF- κ B signaling. *Nat. Commun.* 10 (1), 860. doi:10.1038/s41467-019-08802-0
- Parikh, J. R., Klinger, B. Y. X., Marto, J. A., Blüthgen, N., and Bluthgen, N. (2010). Discovering causal signaling pathways through gene-expression patterns. *Nucleic Acids Res.* 38, W109–W117. doi:10.1093/nar/gkq424
- Peidli, S., Green, T. D., Shen, C., Gross, T., Joseph, M., Taylor-King, J., et al. (2022). scPerturb: Information resource for harmonized single-cell perturbation data. bioRxiv. doi:10.1101/2022.08.20.504663
- Piran, M., Karbalaee, R., Piran, M., Aldahdooh, J., Mirzaie, M., Ansari-Pour, N., et al. (2020). Can we assume the gene expression profile as a proxy for signaling network activity? *Biomolecules* 10 (6), 850. doi:10.3390/biom10060850
- Pizzorno, A., Olivier, T., Nicolas de Lamballerie, C., Julien, T., Padey, B., Traversier, A., et al. (2019). Repurposing of drugs as novel influenza inhibitors from clinical gene expression infection signatures. *Front. Immunol.* 10, 60. doi:10.3389/fimmu.2019.00060
- Preuer, K., Lewis, R. P. I., Hochreiter, S., Bulusu, K. C., and Klambauer, G. (2018). Sepp hochreiter, andreas bender, krishna C. Bulusu, and günter KlambauerDeepSynergy: Predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 34 (9), 1538–1546. doi:10.1093/bioinformatics/btx806
- Pushpakom, S., Iorio, F., Eyers, P. A., Jane Escott, K., Hopper, S., Wells, A., et al. (2019). Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168
- Replogle, J. M., Saunders, R. A., Pogsan, A. N., Hussmann, J. A., Alexander, L., Guna, A., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. *Cell.* 185 (14), 2559–2575.e28. doi:10.1016/j.cell.2022.05.013
- Ritchie, M. E., Phipson, B., Wu, D. H., Yu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Rydenfelt, M., Klinger, B., Klünemann, M., and Blüthgen, N. (2020). SPEED2: Inferring upstream pathway activity from differential gene expression. *Nucleic Acids Res.* 48 (W1), W307–W312. doi:10.1093/nar/gkaa236
- Saez-Rodriguez, J., and Blüthgen, N. (2020). Personalized signaling models for personalized treatments. *Mol. Syst. Biol.* 16 (1), e9042. doi:10.15252/msb.20199042
- Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhrlitz, F., Sauer, S., et al. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* 9 (1), 20. doi:10.1038/s41467-017-02391-6
- Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., et al. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3 (96), 96ra77. doi:10.1126/scitranslmed.3001318
- Smith, I., and Haibe-Kains, B. (2022). Similarity bias from averaging signatures from the connectivity Map. bioRxiv. doi:10.1101/2022.01.24.477615
- Stathias, V., Jermakowicz, A. M., Maloof, M. E., Forlin, M., Walters, W., Suter, R. K., et al. (2018). Drug and disease signature integration identifies synergistic combinations in glioblastoma. *Nat. Commun.* 9 (1), 5315. doi:10.1038/s41467-018-07659-z
- Subramanian, A., Narayan, R., Mcorsello, S., 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 (6), 1437–1452. doi:10.1016/j.cell.2017.10.049
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, Benjamin L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Szalai, B., and Saez-Rodriguez, J. (2020). Why do pathway methods work better than they should? *FEBS Lett.* 594 (24), 4189–4200. doi:10.1002/1873-3468.14011
- Szalai, B., Subramanian, V., Holland, C. H., Alföldi, R., Puskás, L. G., and Saez-Rodriguez, J. (2019). Signatures of cell death and proliferation in perturbation transcriptomics data—from confounding factor to effective prediction. *Nucleic Acids Res.* 47 (19), 10010–10026. doi:10.1093/nar/gkz805
- Trapotsi, M.-A., Hosseini-Gerami, L., and Bender, A. (2022). Computational analyses of mechanism of action (MoA): Data, methods and integration. *RSC Chem. Biol.* 3 (2), 170–200. doi:10.1039/d1cb00069a
- 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 (3), 564–576. doi:10.1016/j.cell.2017.06.010
- Türe, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., et al. (2021). Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* 17 (3), e9923. doi:10.15252/msb.20209923
- van der Meer, D., Barthorpe, S., Yang, W., Howard, L., Hall, C., Gilbert, J., et al. (2019). Cell model passports—A hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res.* 47 (D1), D923–D929. doi:10.1093/nar/gky872
- Wang, W., Bao, J., Zheng, S., Huang, S., Aldahdooh, J., Wang, Y., et al. (2022). A gene essentiality signature for studying the mechanism of action of drugs. bioRxiv. doi:10.1101/2022.11.07.514541
- Wang, Z., Monteiro, C. D., Jagodnik, K. M., Fernandez, N. F., Gundersen, G. W., Rouillard, A. D., et al. (2016). Extraction and analysis of signatures from the gene expression omnibus by the crowd. *Nat. Commun.* 7, 12846. doi:10.1038/ncomms12846

- Woo, H., Shimoni, Y., Subramaniam, P., Iyer, A., Nicoletti, P., Martínez, M. R., et al. (2015). Elucidating compound mechanism of action by network perturbation analysis. *Cell*. 162 (2), 441–451. doi:10.1016/j.cell.2015.05.056
- Yadav, B., Wennerberg, K., Aittokallio, T., and Tang, J. (2015). Searching for drug synergy in complex dose-response landscapes using an interaction potency model. *Comput. Struct. Biotechnol. J.* 13, 504–513. doi:10.1016/j.csbj.2015.09.001
- Yang, M., Jaaks, P., Dry, J., Garnett, M., Menden, M. P., and Saez-Rodriguez, J. (2020a). Stratification and prediction of drug synergy based on target functional similarity. *NPJ Syst. Biol. Appl.* 6 (1), 16. doi:10.1038/s41540-020-0136-x
- Yang, X., Kui, L., Tang, M., Li, D., Wei, K., Chen, W., et al. (2020b). High-throughput transcriptome profiling in drug and biomarker discovery. *Front. Genet.* 11, 19. doi:10.3389/fgene.2020.00019
- 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 (1), 4307. doi:10.1038/s41467-018-06500-x
- Yuan, B., Shen, C., Luna, A., Korkut, A., Debra, S., Ingraham, J., et al. (2021). Marks, John Ingraham, and Chris Sander CellBox: Interpretable machine learning for perturbation biology with application to the design of cancer combination therapy. *Cell. Syst.* 12 (2), 128–140.e4. doi:10.1016/j.cels.2020.11.013
- Zhao, W., Li, J., Chen, M. M., Luo, Y., Ju, Z., Johnson-Camacho, K., et al. (2020). Large-scale characterization of drug responses of clinically relevant proteins in cancer cell lines. *Cancer Cell*. 0 (0), 829–843.e4. doi:10.1016/j.ccell.2020.10.008
- Zheng, M., Okawa, S., Bravo, M., Chen, F., and Martínez-Chantar, M-L. (2022). ChemPert: Mapping between chemical perturbation and transcriptional response for non-cancer cells. *Nucleic Acids Res.* 51, D877–D889. doi:10.1093/nar/gkac862
- Zhong, F., Wu, X., Yang, R., Li, X., Wang, D., Fu, Z., et al. (2022). Drug target inference by mining transcriptional data using a novel graph convolutional network framework. *Protein and Cell*. 13 (4), 281–301. doi:10.1007/s13238-021-00885-0