

REVIEW ARTICLE OPEN



Machine learning approaches to drug response prediction: challenges and recent progress

George Adam^{1,2,3}, Ladislav Rampásek^{1,2,3,4}, Zhaleh Safikhani^{1,3,5}, Petr Smirnov^{1,3,6}, Benjamin Haibe-Kains^{1,2,3,5,6} and Anna Goldenberg^{2,3,4}✉

Cancer is a leading cause of death worldwide. Identifying the best treatment using computational models to personalize drug response prediction holds great promise to improve patient's chances of successful recovery. Unfortunately, the computational task of predicting drug response is very challenging, partially due to the limitations of the available data and partially due to algorithmic shortcomings. The recent advances in deep learning may open a new chapter in the search for computational drug response prediction models and ultimately result in more accurate tools for therapy response. This review provides an overview of the computational challenges and advances in drug response prediction, and focuses on comparing the machine learning techniques to be of utmost practical use for clinicians and machine learning non-experts. The incorporation of new data modalities such as single-cell profiling, along with techniques that rapidly find effective drug combinations will likely be instrumental in improving cancer care.

npj Precision Oncology (2020)4:19; <https://doi.org/10.1038/s41698-020-0122-1>

INTRODUCTION

Cancer is a leading cause of death worldwide and the most important impediment to increasing life expectancy in every country of the world in the 21st century¹. Fortunately, from 2011 to 2015, there has been a small but prominent decrease in death rates for all races/ethnicities combined for 11 out of 18 most common cancers among men and 14 of the 20 most common cancers among women. The continued decreases in death rates for colorectal cancer, prostate cancer and female breast cancer are largely due to advances in early detection and more effective treatments². In this review, we will focus on the computational challenges of identifying the best treatment that improves chances of successful recovery.

Until recently, treatments were chosen based on the type of cancer in a one-size-fits-all manner. We are now witnessing the advent of precision oncology^{3–5} that takes into account patients' genomic makeup for treatment decisions^{3,6,7}. Treatment approval based on tumor-site agnostic molecular aberration biomarkers has become reality. The year 2017 marked the first FDA approval of such a treatment⁸. Based on clinical trials in 15 types of cancer, pembrolizumab was approved for treatment of solid tumors with mismatch repair deficiency or high microsatellite instability⁹. Larotrectinib is another promising treatment, targeting the tropomyosin receptor kinase gene fusion in a variety of cancers¹⁰. Unfortunately, there are no established biomarkers for majority of the anticancer drug compounds. Identification of reliable biomarkers is a challenge not only for the most commonly used cytotoxic drugs, but also in the case of targeted therapies as the drug targets alone are generally poor therapeutic indicators^{11,12}.

Discovery of biomarkers predictive of drug response and development of multivariate companion diagnostics require efficient computational tools and substantial number of samples. Traditional statistical models and more sophisticated machine learning approaches have been used to build predictors of drug

response and resistance both in the clinical¹³ and preclinical¹⁴ settings. As predictive models increase in complexity, the number of observations required to train these models increases as well. While omic profiles and clinical outcomes of patients are the most relevant data sources for the development of clinically relevant predictors, these datasets are often limited in size due to many factors including high costs, limited accrual rates, and complex regulatory landscape. In addition, by the nature of the experiment, unbiased testing of multiple therapeutic strategies for the same patient in the patient itself is practically infeasible. Cancer models provide access to patient tumors in preclinical models, both in vivo and in vitro, allowing researchers to test multiple drugs and combinations in parallel¹⁴. Although these preclinical models recapitulate patient therapy response to varying degrees, they provide massive amounts of pharmacogenomic data for drug response prediction. Here we review the recent applications of machine learning to prediction of response to monotherapies and identification of combination therapies (Fig. 1).

PREDICTION OF RESPONSE TO MONOTHERAPIES

In vitro and ex vivo tumor models

Large-scale efforts to associate molecular profiles with drug response phenotypes in preclinical models date back to the late 90s when the National Cancer Institute Developmental Therapeutics Program released large-scale pharmacogenomic data of 60 cancer cell lines (NCI60) screened with tens of thousands of chemical compounds, including a large panel of FDA-approved drugs¹⁵. NCI60 facilitated several drug discoveries, notably a 26S proteasome inhibitor bortezomib that is now used in multiple myeloma treatment¹⁵. Since then, high-throughput in vitro drug screens of cancer cell lines (CCLs) derived by immortalization of human cancer cells became popular experimental bases for discovery of multi-omic underpinnings of drug sensitivity and

¹Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada. ²Department of Computer Science, University of Toronto, Toronto, ON, Canada. ³Vector Institute, Toronto, ON, Canada. ⁴Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada. ⁵Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada. ⁶Ontario Institute for Cancer Research, Toronto, ON, Canada. ✉email: benjamin.haibe-kains@uhnresearch.ca; anna.goldenberg@utoronto.ca

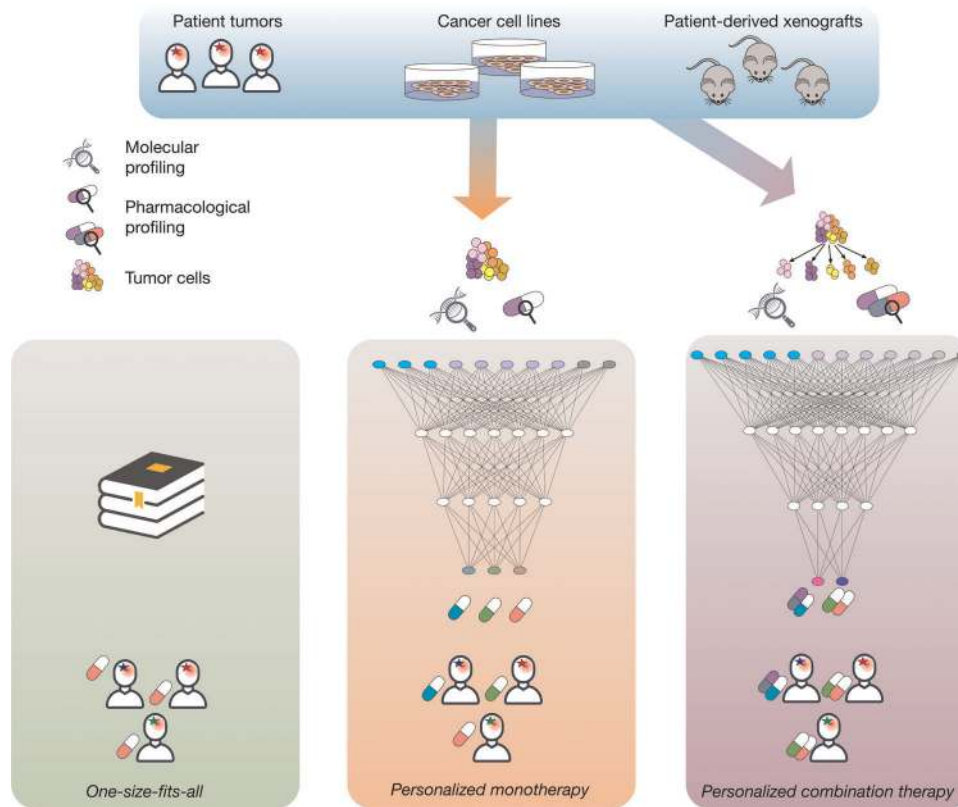


Fig. 1 Graphical abstract. Patient data are limited, so to predict drug response, much of the existing literature use model system data, e.g. immortalized cell lines and PDX. **a** Currently most patients in cancer are still treated in a one-size-fits-all manner according to the type (or subtype) of cancer they have. **b** There is a growing number of examples of personalizing monotherapy in practice, where depending on the mutations in the tumor, the patient can be prescribed a targeted drug. This approach is applicable to fewer than 20% of the patients. The computational contribution is to take a large number of model systems and patients, when available and construct a predictive model to identify the best drug for majority of the patients. **c** Due to tumor heterogeneity and acquired drug resistance, monotherapies may not be effective, there is currently a growing body of work predicting drug synergy and effective drug combinations. Originally these models were trained using bulk data, but more recently, single-cell data-based approaches are starting to show promise. The person symbol in the figure was obtained from dryicons.com. The black magnifying glass is courtesy of Stanislav Tischenko under the Creative Commons Attribution 3.0 License.

resistance¹⁶. Since this seminal study, multiple large-scale databases have been publicly released to the cancer research community^{17,18}. More recently, advances in growing tumors in animal models enabled the generation of large collection of patient-derived xenografts (PDX) to monitor tumor growth with and without drug treatment in mice¹⁹. Novartis published the largest PDX-based pharmacogenomic dataset to date, referred to as the PDX Encyclopedia²⁰. The NCI recently announced the Patient-Derived Models Repository (PDMR) with comprehensive molecular profiling and commitment to release pharmacological profiles in the future. A series of databases and tools have been developed recently to harmonize and make easily available multiple pharmacogenomic studies investigating anticancer monotherapies (Table 1).

Methods for monotherapy prediction

The availability of commercial drug response prediction approaches is limited. In fact, publicly available methods mainly consist of biomarker assays which measure quantities such as gene expression and determine whether or not a specific therapy linked to the biomarker assay would be effective for a given patient. Most of these assays and predictive models are univariate, with only a few multivariate assays that are based on simple statistical and machine learning approaches (the OncotypeDx²¹ and MAMMAPRINT²² models for breast cancer are based on a linear regression model and a nearest centroid model,

respectively). Thus, this review focuses on academic approaches to drug response prediction since they significantly outnumber commercial approaches, are more transparent, and address the more difficult task of predicting the efficacy of multiple drugs without knowing ahead of time the useful features for the task.

The most typical computational approaches to drug response prediction, specifically in preclinical models, consist of (1) quantification of drug response; (2) molecular feature selection or dimensionality reduction of the cellular measurements; (3) machine learning model fitting to predict drug response; and (4) model evaluation^{23,24}. Multiple studies explored which genomic modalities harbor the most predictive signal of drug response by analyzing performance of predictive models. The most commonly utilized modalities include single nucleotide variations, copy number variations, RNA expression, methylation, and proteomics. Despite their widespread use in clinical settings, mutations and copy number variations have been shown to account for only a small subset of candidate biomarkers, while gene expression, methylation and protein abundance are regarded as the most predictive modalities^{25–27}, each can be complemented by the multi-omic view of the cancer^{28–30}. Perhaps the main obstacle in effectively leveraging all data modalities is fusing them while ignoring redundancies. A combined set of measurements can reach hundreds of thousands of features, while the number of available patients or cell lines remains in the hundreds. Such a high feature to sample ratio is bound to lead to overfitting where

Table 1. Platforms harmonizing preclinical pharmacogenomic datasets and providing basic processing functions for biomarker discovery.

Platforms	Cancer models	# Models	# Drugs	URL	References
PharmacoGx PharmacoDB	Cell lines	1691	759	https://bioconductor.org/packages/PharmacoGx/ http://pharmacodb.ca/	17,111
GDSCTools	Cell lines	1001	265	https://gdsc.tools.readthedocs.io	112
CellminerCDB	Cell lines	~1000	~50,000	https://discover.nci.nih.gov/cellminerfdb/	113
CancerDP	Cell lines	1061	24	http://crdd.osdd.net/raghava/cancerdp/index.php	114
PDXFinder	PDX	567	33	https://www.pdxfinder.org/	Unpublished
Xeva	PDX	277	61	https://github.com/bhklab/Xeva	115
Cancer-Drug eXplorer	2D cell cultures	462	60	http://cancerdrugexplorer.org/	116

a model can perfectly fit the limited size training set, yet will have poor generalization performance when tested on new data. This limits the class of applicable predictive models to those with low complexity such as support vector machines or logistic regression since high complexity models like deep neural networks require many samples to avoid overfitting. Successful applications of deep learning in domains such as image classification or machine translation have worked due to a more favorable measurement to sample ratio ($N > D$) in addition to architectures that mimic the human brain and limit overfitting such as convolutional neural networks. Developing neural network architectures with an effective inductive bias for genomics will allow the complex underlying cancer biology to be better modeled compared to linear models which reduce risk of overfitting at the cost of introducing significant modeling bias. Another approach to deal with the limited number of samples typically available in drug response prediction experiments is feature selection. Feature selection removes features such as the gene expression of genes which are determined to be uninformative for the phenotype being predicted. This improves the ratio of features to samples, and a common to feature selection is univariate feature selection where only features highly correlated with the phenotype are kept. Multivariate approaches to feature also exist and consider sets of features at a time since any single feature individually might not be predictive of the outcome, but that does not imply that a collection of features is uninformative as well. Papillon-Cavanagh et al.³¹ identified univariate feature selection as a robust selection approach, later improved by minimum Redundancy, Maximum Relevance (mRMR) Ensemble feature selection³². Costello et al. and Jang et al. performed extensive comparative analyses of machine learning methods for drug response prediction in cancer cell lines, recommending using elastic net or ridge regression with input features from all genomic profiling platforms^{27,29}. Costello et al. summarized a crowdsourced DREAM drug prediction challenge²⁹, revealing two leading trends among the most successful methods. First, the importance of the ability to model nonlinear relationships between data and outcomes, and second, the incorporation of prior knowledge, e.g. biological pathways. The challenge winning model, Bayesian multitask multiple kernel learning method³³, incorporated both of these approaches together with multi-drug learning³⁴. Such multitask framing of the prediction problem is highly effective as it enables a more efficient use of available data when tuning parameters. Specifically, instead of building separate prediction models for each drug thereby using just a subset of the data, a single model trained with all the data that has some parameters shared amongst all the drugs, and some drug-specific parameters is the better choice.

Nonlinear relationships are of utmost importance since many cellular processes follow nonlinear dose-response relationships such as the activation of MAPK via Progesterone in oocytes³⁵. Furthermore, models encoding prior biological knowledge have improved and more stable feature selection since noisy gene-level measurements can be abstracted into gene sets that have been

experimentally validated to be involved in cancer-related processes. Lee et al.³⁶ developed a method that integrates disease relevant multi-omic prior information to prioritize gene-drug associations. Most recently, Zhang et al.³⁷ and Wang et al.³⁸ introduced methods based on similarity network fusion and similarity-regularized matrix factorization, respectively, that take into account similarity among cell lines, drugs and targets. Drug chemical features and similarities were shown to be a promising additional information that can improve drug response prediction performance. There is no canonical way of incorporating drug features into most predictive models since it is difficult to encode how the drug features and omics features interact. Future models that address this shortcoming are likely to outperform competitors that do not, due to the highly informative content of molecular fingerprints. Specifically, a predictive model in a multitask setting can take compounds with known molecular targets, use the similarity computed between the molecular fingerprints, and more effectively tune parameters using similarity between compounds for parameter regularization.

Deep learning methods for monotherapy prediction

The use of neural networks for drug response prediction dates back to the 90s. El-Deredy et al. showed that a neural network trained on tumor nuclear magnetic resonance (NMR) spectra data has potential as a drug response predictor in gliomas, and may be used to provide information about the metabolic pathways involved in drug response³⁹. Neural networks, however, did not become a method of choice for monotherapy prediction yet. In fact, despite the recent prevalence of deep neural network (DNN) methods across many areas and industries, including related fields, such as computational chemistry^{40–45}, DNNs have only fairly recently found their way into the drug response prediction. The reason for this is the typically low ratio of the number of samples to the number of measurements per sample that does not favor traditional feedforward neural architectures. Overparameterization in these models easily leads to overfitting and poor generalization to new datasets. However, in recent years, more public data has become available and newly developed deep neural network models are showing promise. For example, Chang et al.⁴⁶ developed the CDRscan model, featuring a convolutional neural network architecture trained on a dataset of ~1000 drug response experiments per compound. Their model achieved significantly improved performance compared to other classical machine learning approaches such as Random Forests and SVM. Part of why CDRscan performed better than these baseline models resides in its ability to integrate genomic data and molecular fingerprints. In addition, its convolutional architecture has shown to be effective in many machine learning domains. Taking inspiration from already well-established neural architectures, and modifying their structure to properly handle genomic data is certainly a promising future direction.

Table 2. Computational tools for monotherapy prediction.

Name	Availability	Purpose	Methodology and features	Reference
HNMDRP	Matlab and R code	Drug response prediction in CCLs	Genomic and compound features combined with drug–target interaction and PPI	³⁷ Source code: https://github.com/USTC-Hilab/HNMDRP
KRL	Python code	Drug prioritization (ranking) in CCLs transferable to patients	Kernelized rank learning using genomic features, (predominantly gene expression)	¹¹⁷ Source code: https://github.com/BorgwardtLab/Kernelized-Rank-Learning
CDRscan	Web Application ^a	Drug response prediction in CCLs	Deep neural network trained on somatic mutations and drug compound fingerprints	⁴⁶
Dr.VAE	Python code	Drug response prediction in CCLs	Semi-supervised Variational Autoencoder of gene expression that incorporates drug perturbation effects	⁴⁸ Source code: https://github.com/rampasek/DrVAE
CancerDP	Web Application	Drug response prediction in CCLs	SVM models using (combination of) genomic features (mutations, CNVs, expression levels)	¹¹⁴ Webserver: http://crdd.osdd.net/raghava/cancerdp/
BMTMKL	Matlab and R code	Drug response prediction in CCLs	Bayesian multiview (original genomic modalities + aggregated views) multitask model	²⁹ Source code: https://github.com/mehmetgonen/bmtmk1

A non-exhaustive summary of the most recent monotherapy prediction methods with an available web service or source code.

^aA web application has been promised by the authors, but no official implementation yet as of February 2020.

Another promising direction is autoencoders that are able to learn from smaller datasets. An autoencoder is a neural network that compresses its input and tries to reconstruct the original data from the compressed representation. This is quite useful for feature extraction as shown by Way and Greene⁴⁷ where a 5000 dimensional gene expression profile was compressed into just 100 dimensions, some of which represented phenotypically relevant features such as patient sex or melanoma status. Rampásek et al.⁴⁸ evaluated semi-supervised variational autoencoders on monotherapy response prediction and developed an extension—a joint drug response prediction model, Dr.VAE, that leveraged pre- and post-treatment gene expression in cell lines, showing improved performance in drug response prediction on a variety of FDA-approved drugs compared comprehensively to many classical machine learning approaches. This improvement could potentially have been even greater if the model was setup in a multitask fashion in combination with molecular fingerprints. Dincer et al.⁴⁹ developed DeepProfile, a method that combined variational autoencoders to learn 8-dimensional representation of gene expression in AML patients and then used this representation to fit a Lasso linear model for drug response prediction with improved performance compared to no feature extraction. Similarly, Chiu et al.⁵⁰ pretrained autoencoders on mutation data and expression features on TCGA dataset and subsequently trained a deep drug response predictor. What differentiates their method from others is the use of pretraining. Pretraining allows for using unlabeled data from other sources such as TCGA, instead of just the gene expression profiles available from the drug response experiments, thereby significantly increasing the number of samples available and improving performance compared to using just the labeled data. The brief summary of methods is available in Table 2. The trend of model development shows that as more data become available and deep learning methods become better adapted to high dimensional/low sample size data, there is hope for convergence and creation of sophisticated models that will likely push the field of computational drug response prediction forward to eventually become clinically relevant.

RESISTANCE TO MONOTHERAPY

While drug response prediction can help pick an optimal therapy given the current molecular characteristics of the cancer cells, tumors often exhibit drug resistance over the course of the treatment. Consequently, patients that respond initially to therapy regress as their cancer either adapts to overcome the chosen

treatment, or an existing resistant subclone repopulates the tumor⁵¹. Understanding the common mechanisms cancers use to develop resistance can help inform treatment approaches to counteract this phenomena.

For therapies inhibiting the activity or signaling of their target, a common mechanism towards resistance is feedback selecting for upregulated expression of the target protein. For example, resistance to 5-FU has been demonstrated to arise from the amplification of its target thymidylate synthase (TS)⁵², with corresponding overproduction of TS enzyme and mRNA transcripts⁵³. Furthermore, especially for tyrosine kinase inhibitors, tumors will evolve to re-activate pathways downstream of the targeted protein. A classical example is the resistance to the EGFR inhibitor Gefitinib which can often be explained by an acquired T790M mutation reducing drug binding affinity⁵⁴.

For DNA damaging compounds or compounds inhibiting DNA repair, altered DNA damage response can lead to resistance. Studies have shown that treatment with cisplatin, a DNA damaging agent usually effective against BRCA deficient cancers, can lead to mutations restoring BRCA function and subsequently the activity of the Homologous Repair (HR) pathway^{55,56}. Furthermore, studies suggest that secondary alterations to DNA damage response proteins can shift the response from the error-prone Non-Homologous End Joining pathway to HR, reducing sensitivity to DNA damaging agents⁵⁷. Other mechanisms of resistance include modifications to enzymes involved in drug metabolism to either reduce conversion of drugs to active forms or deactivate the compound^{58,59}, and more recently, intra-tumor heterogeneity (ITH)⁶⁰. As this review focuses on drug response prediction, not enough depth is provided to discussing how tumors acquire resistance to therapies, or how therapies work. Readers are referred to work by Holohan et al.⁵¹, Housman et al.⁵⁹, and Malhotra and Perry⁶¹ for a comprehensive discussion on this topic. For more details on the biological complexity of cancer in general, readers are referred to the review articles by Blackadar⁶², and Bertram⁶³.

COMBINATION THERAPIES

Drug combinations are crucial for addressing the issue of drug resistance and preventing recurrence caused by a negligible amount of remaining cancer cells. Synergistic combinations can also reduce toxicity by allowing for lower doses of either drug to be used. By enabling reduced doses, drug combinations can further increase the feasibility of drug repurposing by increasing

Table 3. Methods to infer tumor clonal composition from bulk DNA sequencing data.

Name	Using SSM or CNV for phylogeny reconstruction	Joint Deconvolution and Phylogeny inference?	Reference
PhyloWGS	SSM and percolated CNV mixing proportion estimates	Joint Inference	118
Canopy	Both	Joint Inference	119
SPRUCE	SSM and percolated CNV mixing proportion estimates	Joint Inference	120
PASTRI	SSM only	Two step clustering and Phylogeny Inference	121
PyClone	SSM only, corrects VAFs for CNV, does not use in reconstruction explicitly	Clustering and Identifying clonal genotypes only	122
SciClone	SSM only	Clustering and Identifying clonal genotypes only	123
THetA2	CNV only	Clustering and Identifying clonal genotypes only	124

the potency of compounds that are only effective at clinically dangerous doses⁶⁴.

Trial and error combination design has limited applicability in the clinic due to time constraints and potential hazardous exposure to toxic combinations without improving efficacy. For example, Hecht et al.⁶⁵ performed a clinical trial for metastatic colorectal cancer (mCRC) patients involving the targeted compound bevacizumab, either oxaliplatin or irinotecan as a chemotherapeutic agent, and an optional addition of a human antibody panitumumab. The purpose of the trial was to evaluate benefit conferred by panitumumab. It was revealed that for the cohort that used oxaliplatin as a chemotherapeutic agent, survival was 5 months lower for patients that also received panitumumab, and there was a significant increase in adverse effects such as infections and pulmonary embolism compared to patients that did not receive panitumumab. Tol et al.⁶⁶ also performed a clinical trial for mCRC, using combination of capecitabine, oxaliplatin, and bevacizumab, as the baseline treatment to investigate cetuximab. Patients that received cetuximab had a shorter progression-free survival and reported significantly more adverse effects compared to patients that did not receive cetuximab.

One promising direction for a setting where the goal is to study a constrained set of options to design an optimal treatment plan for a patient is adaptive trials via reinforcement learning⁶⁷. The probabilistic ranking given by their method potentially allows for identifying when tumors develop drug resistance by analyzing when drug combinations are given priority over individual treatments. While this work, performed on PDX, learns more complex yet more effective policies in terms of survival than currently offered in the clinic, it is not clear how to mitigate the potential risks of exploration needed for reinforcement learning. We do hope that this direction is given its due consideration in the clinic since these early results appear to be very promising.

The limits of trial and error in the clinic can also be overcome in vitro with the use of preclinical models in the form of immortalized cancer cell lines or cell lines derived from patient biopsies. Patient-derived cancer models allow screening drug combinations in parallel without subjecting patients to serious toxicity risk (Table 4). Unfortunately, due to the sheer number of possible drug combinations, it is not possible to explore their potential antagonism, additive or synergistic effects⁶⁸, so there is a need for methods that can predict combination therapy response prior to experimentally validating it.

Methods for combination therapy prediction

Many computational methods have been developed to predict anticancer drug combination synergy based on a variety of genomic, drug structure, and biological network data. These methods vary in how much drug combination screening data is required, if used at all. Drug combination screening data refers to testing cancer models with combinations of two or more drugs rather than a single drug. A typical combination experiment setup

involves testing two drugs at 8 different half-log dilution concentrations each including the null concentration as a control⁶⁹. This gives rise to an 8x8 dose-response matrix. Using a 384-well assay plate, six pairs of drugs can be screened at once in this arrangement. Once cells are incubated in the wells for a sufficient amount of time, usually 72 h, a cell viability readout is conducted to determine the number of viable cells in each well. The collected data is then processed using a tool such as SynergyFinder⁷⁰ to quantify the drug combination response compared to individual compound response based on a variety of models. As an example, the Bliss independence model⁷¹ provides a score under the assumption that the two drugs act independently, so measurement above this score indicates synergy. For more details on different synergy scores as well as experimental design of drug combination studies, the reader is referred to the experimental design guide by He et al.⁶⁹. The number of experiments increases exponentially with the number of drugs tested in combination, making these combination screens both logistically complex and expensive. It is therefore favorable to have a method which does not require significant amounts of combination screens. Several approaches for drug synergy prediction described in the literature instead use a combination of either perturbation experiments or sensitivity experiments coupled with drug target and drug structure data. For example, the work done by Li et al.⁷² leverages gene expression perturbation data, measured as the difference in gene expression before and after treatment, to compute various statistics about differentially expressed genes as the main pharmacogenomic features. Additionally, the authors extracted drug physicochemical properties, distance between drug targets in PPI networks, and Jaccard similarity between targeted pathways to represent biological and chemical prior knowledge. These features were then used to train a random forest model to perform the binary prediction task of whether a drug combination is synergistic or not. Gayvert et al.⁷³ also made predictions with random forests by using both single-drug response values and combination therapy response values when available. Interestingly, they did not leverage drug structure information nor gene expression profiles when making predictions. This is a drawback since drug structure information is easily available, and including it may improve performance, but it provides flexibility in not having to measure gene expression. However, their framework is broadly applicable, and their results indicate that even a small number of drug combination experiments can have a great performance benefit when used to train a model that makes predictions using primarily single-drug response data.

There is a class of drug combination optimization approaches that interacts with the user by suggesting promising combinations to test. Both Weiss et al.⁷⁴ and Nowak-Sliwinska et al.⁷⁵ use Feedback System Control (FSC) to iteratively refine drug combinations and suggest new ones to test in vitro. The process works by first starting with some randomly selected drug combinations for some range of doses. This group of

Table 4. Drug combination datasets.

Dataset Name	Type	# Combinations	# Drugs	# Patients/ cell Lines	URL	Ref
Drug Combination Database	Clinical	1363	904	~140,000	http://www.cls.zju.edu.cn/dcdb/	125
Merck	In vitro	583	38	39	http://mct.aacrjournals.org/highwire/filestream/53222/field_highwire_adjunct_files/3/156849_1_supp_1_w2lrww.xls	126
AstraZeneca-Sanger Drug Combination Dataset	In vitro	910	118	85	https://www.synapse.org/#!Synapse:syn4231880/wiki/235645	30
NCI ALMANAC	In vitro	5,000+	105	60	https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC	78

combinations is then mutated using Differential Evolution (DE) to propose new drug combinations that are to be tested in vitro, and whose efficacy will be compared against the original randomly selected combinations. For each mutated combination, if that combination had higher efficacy than the original random combination that it was created from, then the new combination is kept, otherwise the original combination is kept. This procedure is repeated until some convergence criterion is met. This approach seems to be very effective in practice because the efficacy versus drug combination surface is smooth thereby allowing FSC to converge in 10–15 iterations. Lastly, the optimal drug combination identified by DE and evaluated in vitro is further optimized to eliminate redundant compounds or compounds having an antagonistic effects. Importantly, FSC based approaches are not limited in the number of drugs used in a given combination, unlike many methods that are created specifically for pairs of drugs. It might be possible to accelerate the convergence of FSC methods by including genomic or chemical data since both methods described above perform the optimization without considering drug targets or drug similarities.

Deep learning methods for combination therapy prediction

The most extreme prediction scenario is to not use drug response data at all when building a model. This is done by Preuer et al.⁷⁶ where the authors only leverage transcriptomic data and drug structure data to predict Loewe score which quantifies the excess over the expected response if the two drugs used in a combination were the same compound. What further differentiates this work from previous works is that the authors use deep learning to achieve state-of-the-art performance compared to baseline models such as gradient boosting machines, random forests, and support vector machines. Xia et al.⁷⁷ used deep learning as a means of simultaneously extracting and integrating features from multiple data types to predict the efficacy of drug pairs. Combination response data as well as gene expression, microRNA, and protein abundance from the NCI-ALMANAC dataset was used⁷⁸. Additionally, drug features were obtained using Dragon software⁷⁹ which provides chemical fingerprints and other properties. Each data type was passed through its own submodel where a submodel is just a deep fully connected neural network in order to obtain useful features and perform dimensionality reduction. Then, these features for the different data types were concatenated and passed through a final submodel that uses residual connections in order to predict the drug combination score. Ultimately, the authors were able to obtain impressive results with R^2 of 0.92, and much of that explained variance was due to the drug descriptors. These approaches reinforce the importance of newer deep learning methods such as molecular graph convolution to extract task-specific molecular fingerprints. A summary of tools related to drug combinations is provided in Table 5. In terms of the availability, there are more synergy visualization tools rather than synergy

prediction tools available to date. We hope that this trend will change as more researchers work on this important area and provide their tools in publicly available packages.

DRUG COMBINATION DISCOVERY USING SINGLE-CELL SEQUENCING

The development of single-cell sequencing technologies has given researchers a new set of tools to interrogate tumor heterogeneity. Single-cell DNA sequencing (scDNAseq), can be used to more directly investigate the clonal structure of a tumor. It works by isolating individual cells and performing whole genome amplification to increase the amount of DNA present in order to be detectable by a DNA sequencer⁸⁰. These data can be used to directly reconstruct the unique genotypes as well as to estimate the clonal fraction within the sample. Bulk DNA sequencing does not have these abilities, so simply identifying populations of cells with different mutations can already significantly improve treatment plans (Table 3). However, scDNAseq data suffers from increased noise—each cell has only two copies of each genomic locus, requiring amplification before sequencing⁸¹. The amplification process can introduce errors into the sequenced reads, and amplification can be uneven across the genome as well as between cells, introducing bias into the observed reads. Computational approaches estimating tumor clonal composition while taking into account these sources of error have been developed^{82–84}. For a thorough discussion of the methods used to analyze scDNAseq data, we refer the reader to the work by Qi et al.⁸⁵. Interestingly, single-cell RNA sequencing (scRNAseq) is starting to be used to design novel drug combinations through identifying druggable subclones^{86,87}. Unlike DNA, where each cell contains only one copy of each allele (to a total of 6 pg of DNA), there is approximately 30 pg of RNA in a single cell. With the advent of the Chromium platform it is also now possible to sequence the RNA across 100,000s of cells in a single experimental run⁸⁸. Predictive models of drug response could be developed and trained using high-throughput preclinical pharmacogenomic data, and an optimization framework to predict the most efficient and the least toxic combination treatment could be established.

One of the first analyses to examine the influence of treatment on the transcriptome of cancer cells at single-cell resolution was conducted by Suzuki et al.⁸⁹. They first performed single-cell sequencing on four different cell lines derived from lung adenocarcinoma to compare the relative divergence in their gene expression profiles. Even though the average gene expression levels were generally similar, the relative divergences between cell types were pronounced. To investigate how targeted therapy affects individual cells, they treated LC2/ad cell line and the derived resistant version of it with vandetanib, a multi-tyrosine kinase inhibitor. The comparison of single-cell profiles of treated cells versus parental cells identified a wide variety of genes overexpressed by drug stimulation. Particularly in case of LC2/ad, the diversity level of gene abundances between cells was

Table 5. Tools for visualizing, evaluating, and predicting synergistic drug combinations.

Name	Implementation	Purpose	Features	URL
SynergyFinder	Web Application	Evaluating Combo Efficacy	Has 4 different drug interactivity models Computes single-agent effects Computes synergy scores	https://synergyfinder.fimm.fi/
Combeneft	Desktop Application	Evaluating Combo Efficacy	Has 3 different drug interactivity models Meant to handle large batch experiments	https://www.cruk.cam.ac.uk/research-groups/jodrell-group/combeneft
Clmbinator	Web Application	Evaluating Combo Efficacy	Has 1 drug interactivity model	http://cimbulator.bioinfo.cnio.es/CombinationIndex
DIGREM	Web Application	Evaluating Combo Efficacy	Models response curve and gene expression changes after treatment	http://lce.biohpc.swmed.edu/drugcombination/
RACS	R Package	In-Silico Synergy Prediction	Leverages drug target networks and transcriptomic profiles	https://github.com/DrugCombination/RACS
DeepSynergy	Web Application	Predicts Synergy Scores	Selects novel synergistic drug combinations	http://www.bioinf.jku.at/software/DeepSynergy/

significantly reduced after treatment so authors hypothesized that cells lose diversity in response to treatment. Interestingly, target genes of vandetanib, EGFR and RET, were not as affected by the treatment as some of the other off-target genes possibly due to the rigid transcriptional controls over these targets.

Kim et al.⁹⁰ sequenced the transcriptome at single-cell resolution of a primary renal cell carcinoma (pRCC) and its lung metastasis (mRCC) from a patient and paired PDX models to design a combination therapy that would address the heterogeneous nature of the tumor. Whole exome sequencing of the metastatic sample and its PDX model indicated the preservation of major tumor features in the PDX model. In order to predict single-cell response of the RCC to the clinically approved drugs, activity of drug target pathways was estimated by conducting gene set enrichment analysis. Subsequently, cell lines derived from the PDX models were screened with the drugs. Predictive drug response models, based on ridge regression, were built using expression profiles of cancer cell lines from a publicly available drug screening dataset^{91,92} to predict response to the drugs. Authors used ComBat to remove the technical variation between the cell line dataset used for training, the drug response predictors, and single-cell RNA-seq data. Predicted drug response values were substantially correlated with measured sensitivity values (0.65). Accordingly, by considering high sensitivity prediction of cells to Afatinib and Dasatinib and mutually exclusive patterns in the activation status of their signaling pathways in cells, the authors suggested a combination of these two compounds as an efficient therapeutic strategy. In vitro validation in 2D and 3D cultured mRCC cells and in vivo validation in subcutaneous xenografts validated the expected additive effect of the drug combination over monotherapy responses. The administration of this combinatorial therapy is inducing superior growth inhibition by co-targeting mutually exclusive EGFR and Src signaling pathways.

One of the major weaknesses of Kim et al.'s⁹⁰ work is the low number of single cells sequenced. The captured cells may not reflect the true clonality of the patient tumor and might even lead to false discoveries. Recent technological advances in single-cell sequencing made it feasible to capture large numbers of single cells in one experiment. New computational pipelines and approaches have been developed to improve all the steps in processing of the single-cell sequencing data^{93,94}, including tackling noise and dropout in these experiments, normalization techniques, dimensionality reduction^{95–97} and clustering approaches^{98,99}. These rapidly evolving methodologies provide remarkable opportunities for the discovery of biomarkers, prediction of efficient therapies, and the study of mechanisms of acquiring resistance to treatments.

Anchang et al.¹⁰⁰ were the first to use single-cell perturbation experiments to optimize drug combinations. Their model DRUG-NEM required the specification of lineage, intracellular communication, and apoptosis markers that were measured in drug perturbation experiments using Mass Cytometry Time-of-Flight (CyTOF). The objective of the model is to select the minimum number of drugs that creates the maximum perturbation effect on the markers of interest using perturbation data from single-drug experiments. Drug effects were measured using a Bayesian linear model to compute the probability that an intracellular communication marker is differentially expressed between treatment and control. A graphical model is then created from these probabilities using a nested effects model, and all the possible drug combinations are ranked. This approach is limited by having to know ahead of time which markers to use, and this in turn requires knowing the mechanisms of action for the drugs, which in many cases is not available. Nevertheless, this direction for drug response prediction is very promising and will be greatly aided by the burgeoning single cell and drug clonality research.

OPPORTUNITIES AND CHALLENGES: DATA AND DEEP LEARNING

The only standardized metric to date for cancer response is RECIST, and it relies on imaging data, mainly CT and MRI, to determine how tumors grow or shrink in patients. RECIST can handle up to 10 lesions in the patient, prioritized based on the largest lesions, and uses the sum of the lesion diameters (LD) when first measured as the baseline value. In subsequent scans, response is categorized into 4 different categories based on how much the sum of LDs has changed: complete response, partial response, stable disease, and progressive disease. There is no such international standard used to measure response for in vitro preclinical models and RECIST is usually not used in in vivo preclinical models due to costs, thus prohibiting fair comparisons between response prediction methods. Furthermore, some drug response prediction studies frame the task as regression where continuous values such as IC50 are predicted, and others frame the task as classification where a binary value which indicates inhibition or growth is predicted. Reproducibility between cell line based drug response studies remains a challenge due to differences in viability assays, drug concentrations, and cell seeding density¹⁰¹. There is also a need for better data sharing as technical replicates are necessary for estimating within-study variability, yet are sometimes not publicly released¹⁰². Additionally, the studies use a variety of datasets, thereby making quantitative comparisons even less feasible. Instead, qualitative comparisons are made between the methods that consider data

requirements, generalization ability, and capacity to model complex biological interactions and chemical interactions. These comparisons are of great practical use as they provide context and scenarios in which one method is likely better than another.

The success of deep learning across scientific fields followed the collection of large standardized datasets. An additional factor important to broad utilization of deep learning was the growth in available computational power for training these models. Similarly, successful applications of deep learning in predictive oncology followed the growth of high-throughput preclinical datasets. This suggests that with additional data from studies that are more reproducible, deep learning could provide significant improvements over traditional machine learning methods in drug response prediction and drug combination prioritization. Specifically, the end-to-end nature of deep learning allows for extremely effective feature extraction and also enables the integration of multiple distinct data modalities. Additionally, encoding prior biological knowledge in neural networks can be done via several mechanisms such as graph-convolution networks¹⁰³, or conditional scaling which allows for multiplicative relations between features such as a mutation being required for gene expression levels to be relevant. The nonlinear nature of deep neural networks, combined with their inductive bias that allows them to generalize even though they have many more parameters than samples, suggests that promising applications are possible in pharmacogenomics where complex correlation structures exist among features and between features and labels. For example, graph convolutional networks are a promising new way of encoding structural information from molecular graphs¹⁰⁴ and can give application-specific chemical fingerprints that are more specialized for drug response or combination therapy discovery. Another fruitful direction is the use of transfer learning to leverage an abundance of omics data already available. The main obstacle for transfer learning is the large discrepancies between the techniques and experimental protocols used for different studies which lead to batch effects that violate the assumptions on which deep learning relies to generalize to new datasets. The creation of domain adaptation techniques, similar to computer vision¹⁰⁵, specific for omics data will be of immense help in enabling transfer learning. Still, creating architectures with an effective inductive bias for processing omics data is difficult since it is not possible to just rely on the human brain for inspiration like in image analysis. Thus, neural architecture search techniques which remove humans from the design loop by automating the creation and testing of architectures are of key importance in making deep learning more successful in drug response prediction¹⁰⁶. It has recently been shown that the success of architecture search techniques depends significantly on careful design of the search space¹⁰⁷. This requires encoding prior knowledge about potentially effective architecture choices which is certainly less difficult than specifying an entire architecture, but still remains challenging. Deep learning can certainly help in better understanding cancer biology by predicting binding sites or discovering new biomarkers by analyzing RNA transcripts^{47,108,109}. In fact, deep learning has also been used to predict protein-protein interactions¹⁰⁹ which are of increasing interest as potential targets for cancer therapies¹¹⁰, so deep learning will have an impact on both drug discovery and drug response prediction.

The problem of predicting the optimal treatment or combination of treatments for a cancer patient remains unsolved. The approaches reviewed above seek to bring recent advances in machine learning to bear on this challenge, leveraging the growing high-throughput preclinical screening data and new technologies allowing the profiling of tumors on a single-cell level. Promising results in this area should encourage both the investigators working on developing cheaper and more precise high-throughput screens to enable further data collection as well as ML method developers to develop novel tools incorporating

peculiarities of cancer biology. While there remains much work to be done, the field is nascent and offers a path to a truly personalized approach to oncology.

Received: 23 August 2019; Accepted: 17 April 2020;
Published online: 15 June 2020

REFERENCES

- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* <https://doi.org/10.3322/caac.21492> (2018).
- Cronin, K. A. et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer* **124**, 2785–2800 (2018).
- Garraway, L. A., Verweij, J. & Ballman, K. V. Precision oncology: an overview. *J. Clin. Oncol.* **31**, 1803–1805 (2013).
- Doherty, M., Metcalfe, T., Guardino, E., Peters, E. & Ramage, L. Precision medicine and oncology: an overview of the opportunities presented by next-generation sequencing and big data and the challenges posed to conventional drug development and regulatory approval pathways. *Ann. Oncol.* **27**, 1644–1646 (2016).
- Heymach, J. et al. Clinical Cancer Advances 2018: annual report on progress against cancer from the American Society of Clinical Oncology. *J. Clin. Oncol.* **36**, 1020–1044 (2018).
- Twomey, J. D., Brahme, N. N. & Zhang, B. Drug-biomarker co-development in oncology—20 years and counting. *Drug Resist. Updat* **30**, 48–62 (2017).
- Johnson, A. et al. The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. *Drug Discov. Today* **20**, 1433–1438 (2015).
- Prasad, V., Kaestner, V. & Mailankody, S. Cancer drugs approved based on biomarkers and not tumor type—FDA approval of pembrolizumab for mismatch repair-deficient solid cancers. *JAMA Oncol.* **4**, 157–158 (2018).
- Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
- Drilon, A. et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *N. Engl. J. Med.* **378**, 731–739 (2018).
- De Roock, W., De Vriendt, V., Normanno, N., Ciardiello, F. & Tejpar, S. KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol.* **12**, 594–603 (2011).
- Ding, M. Q., Chen, L., Cooper, G. F., Young, J. D. & Lu, X. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. *Mol. Cancer Res.* **16**, 269–278 (2018).
- Perez-Gracia, J. L. et al. Strategies to design clinical studies to identify predictive biomarkers in cancer research. *Cancer Treat. Rev.* **53**, 79–97 (2017).
- Dhandapani, M. & Goldman, A. Preclinical cancer models and biomarkers for drug development: new technologies and emerging tools. *J. Mol. Biomark. Diagn.* **8**, 356 (2017).
- Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6**, 813–823 (2006).
- Macarron, R. et al. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10**, 188–195 (2011).
- Smirnov, P. et al. PharmacDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Res.* **46**, D994–D1002 (2018).
- Ling, A., Gruener, R. F., Fessler, J. & Huang, R. S. More than fishing for a cure: the promises and pitfalls of high throughput cancer cell line screens. *Pharmacol. Ther.* <https://doi.org/10.1016/j.pharmthera.2018.06.014> (2018).
- Aparicio, S., Hidalgo, M. & Kung, A. L. Examining the utility of patient-derived xenograft mouse models. *Nat. Rev. Cancer* **15**, 311–316 (2015).
- Gao, H. et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015).
- McVeigh, T. P. et al. The impact of Oncotype DX testing on breast cancer management and chemotherapy prescribing patterns in a tertiary referral centre. *Eur. J. Cancer* **50**, 2763–2770 (2014).
- Slodkowska, E. A. & Ross, J. S. MammaPrint™ 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev. Mol. Diagn.* **9**, 417–422 (2009).
- Azuaje, F. Computational models for predicting drug responses in cancer research. *Brief. Bioinform.* **18**, 820–829 (2017).
- De Niz, C., Rahman, R., Zhao, X. & Pal, R. Algorithms for drug sensitivity prediction. *Algorithms* **9**, 77 (2016).
- Iorio, F. et al. A landscape of pharmacogenomic interactions in cell. *Cell* **166**, 740–754 (2016).

26. Safikhani, Z. et al. Gene isoforms as expression-based biomarkers predictive of drug response in vitro. *Nat. Commun.* **8**, 1126 (2017).
27. Jang, I. S., Neto, E. C., Guinney, J., Friend, S. H. & Margolin, A. A. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac. Symp. Biocomput.* **19**, 63–74 (2014).
28. Stetson, L. C., Pearl, T., Chen, Y. & Barnholtz-Sloan, J. S. Computational identification of multi-omic correlates of anticancer therapeutic response. *BMC Genomics* **15**(Suppl. 7), S2 (2014).
29. Costello, J. C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).
30. Menden, M. P. et al. A cancer pharmacogenomic screen powering crowd-sourced advancement of drug combination prediction. *bioRxiv* 200451. <https://doi.org/10.1101/200451> (2018).
31. Papillon-Cavanagh, S. et al. Comparison and validation of genomic predictors for anticancer drug sensitivity. *J. Am. Med. Inform. Assoc.* **20**, 597–602 (2013).
32. De Jay, N. et al. mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* **29**, 2365–2368 (2013).
33. Gönen, M. & Margolin, A. A. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* **30**, i556–i563 (2014).
34. Ammad-Ud-Din, M., Khan, S. A., Wennerberg, K. & Aittokallio, T. Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. *Bioinformatics* **33**, i359–i368 (2017).
35. Andersen, M. E., Yang, R. S. H., French, C. T., Chubb, L. S. & Dennison, J. E. Molecular circuits, biological switches, and nonlinear dose–response relationships. *Environ. Health Perspect.* **110**(Suppl. 6), 971–978 (2002).
36. Lee, S.-I. et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat. Commun.* **9**, 42 (2018).
37. Zhang, F., Wang, M., Xi, J., Yang, J. & Li, A. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. *Sci. Rep.* **8**, 3355 (2018).
38. Wang, L., Li, X., Zhang, L. & Gao, Q. Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer* **17**, 513 (2017).
39. El-Deredy, W. et al. Pretreatment prediction of the chemotherapeutic response of human glioma cell cultures using nuclear magnetic resonance spectroscopy and artificial neural networks. *Cancer Res.* **57**, 4196–4199 (1997).
40. Dahl, G. E., Jaitly, N. & Salakhutdinov, R. Multi-task neural networks for QSAR predictions. Preprint at <https://arxiv.org/abs/1406.1231> (2014).
41. Unterthiner, T. et al. Deep learning as an opportunity in virtual screening. in *Proc. Deep Learning Workshop at NIPS, NeurIPS workshop*, Vol. 27, 1–9 (2014).
42. Aliper, A. et al. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* **13**, 2524–2530 (2016).
43. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for Quantum chemistry. in *Proceedings of the 34th International Conference on Machine Learning - Vol. 70*, 1263–1272 (JMLR.org, 2017).
44. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
45. Menden, M. P. et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE* **8**, e61318 (2013).
46. Chang, Y. et al. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci. Rep.* **8**, 8857 (2018).
47. Way, G. P. & Greene, C. S. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. *Pac. Symp. Biocomput* **23**, 80–91 (2018).
48. Rampásek, L. et al. Improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btz158> (2019).
49. Dincer, A. B., Celik, S., Hiranuma, N. & Lee, S.-I. DeepProfile: deep learning of cancer molecular profiles for precision medicine. *bioRxiv* 278739. <https://doi.org/10.1101/278739> (2018).
50. Chiu, Y.-C. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genomics* **12**, 119 (2019).
51. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **13**, 714–726 (2013).
52. Jenh, C. H., Geyer, P. K., Baskin, F. & Johnson, L. F. Thymidylate synthase gene amplification in fluorodeoxyuridine-resistant mouse cell lines. *Mol. Pharmacol.* **28**, 80–85 (1985).
53. Berger, S. H., Jenh, C. H., Johnson, L. F. & Berger, F. G. Thymidylate synthase overproduction and gene amplification in fluorodeoxyuridine-resistant human cells. *Mol. Pharmacol.* **28**, 461–467 (1985).
54. Kobayashi, S. et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
55. Sakai, W. et al. Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451**, 1116–1120 (2008).
56. Edwards, S. L. et al. Resistance to therapy caused by intragenic deletion in BRCA2. *Nature* **451**, 1111–1115 (2008).
57. Bouwman, P. & Jonkers, J. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. *Nat. Rev. Cancer* **12**, 587–598 (2012).
58. Meijer, C. et al. Relationship of cellular glutathione to the cytotoxicity and resistance of seven platinum compounds. *Cancer Res.* **52**, 6885–6889 (1992).
59. Housman, G. et al. Drug resistance in cancer: an overview. *Cancers* **6**, 1769–1792 (2014).
60. Sun, X.-X. & Yu, Q. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacol. Sin.* **36**, 1219–1227 (2015).
61. Malhotra, V. & Perry, M. C. Classical chemotherapy: mechanisms, toxicities and the therapeutic window. *Cancer Biol. Ther.* **2**, 1–3 (2003).
62. Blackadar, C. B. Historical review of the causes of cancer. *World J. Clin. Oncol.* **7**, 54–86 (2016).
63. Bertram, J. S. The molecular biology of cancer. *Mol. Asp. Med.* **21**, 167–223 (2000).
64. Sun, W., Sanderson, P. E. & Zheng, W. Drug combination therapy increases successful drug repositioning. *Drug Discov. Today* **21**, 1189–1195 (2016).
65. Hecht, J. R. et al. A randomized phase III trial of chemotherapy, bevacizumab, and panitumumab compared with chemotherapy and bevacizumab alone for metastatic colorectal cancer. *J. Clin. Oncol.* **27**, 672–680 (2009).
66. Tol, J. et al. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. *N. Engl. J. Med.* **360**, 563–572 (2009).
67. Durand, A. et al. Contextual bandits for adapting treatment in a mouse model of de novo carcinogenesis. In *Proc. 3rd Machine Learning for Healthcare Conference* (eds. Doshi-Velez, F. et al.) Vol. 85, 67–82 (PMLR, 2018).
68. Rationalizing combination therapies. *Nat. Med.* **23**, 1113 (2017).
69. He, L. et al. Methods for high-throughput drug combination screening and synergy scoring. *Methods Mol. Biol.* **1711**, 351–398 (2018).
70. lanevski, A., He, L., Aittokallio, T. & Tang, J. SynergyFinder: a web application for analyzing drug combination dose–response matrix data. *Bioinformatics* **33**, 2413–2415 (2017).
71. Bliss, C. I. The toxicity of poisons applied jointly 1. *Ann. Appl. Biol.* **26**, 585–615 (1939).
72. Li, X. et al. Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif. Intell. Med.* **83**, 35–43 (2017).
73. Gayvert, K. M. et al. A computational approach for identifying synergistic drug combinations. *PLoS Comput. Biol.* **13**, e1005308 (2017).
74. Weiss, A. et al. Rapid optimization of drug combinations for the optimal angiostatic treatment of cancer. *Angiogenesis* **18**, 233–244 (2015).
75. Nowak-Sliwinska, P. et al. Optimization of drug combinations using Feedback System Control. *Nat. Protoc.* **11**, 302–315 (2016).
76. Preuer, K. et al. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **34**, 1538–1546 (2018).
77. Fangfang Xia et al. Predicting tumor cell line response to drug pairs with deep learning. In *Computational Approaches for Cancer Workshop at SC17*. Available at: <http://www.scworkshops.net/cancer2017/> (2017). (Accessed 20 Nov 2018).
78. Holbeck, S. L. et al. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res.* **77**, 3564–3576 (2017).
79. Mauri, A., Consonni, V., Pavan, M. & Todeschini, R. Dragon software: an easy approach to molecular descriptor calculations. *Match* **56**, 237–248 (2006).
80. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).
81. Ortega, M. A. et al. Using single-cell multiple omics approaches to resolve tumor heterogeneity. *Clin. Transl. Med.* **6**, 46 (2017).
82. Ross, E. M. & Markowitz, F. OncoNEM: inferring tumor evolution from single-cell sequencing data. *Genome Biol.* **17**, 69 (2016).
83. Jahn, K., Kuipers, J. & Beerenwinkel, N. Tree inference for single-cell data. *Genome Biol.* **17**, 86 (2016).
84. Roth, A. et al. Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat. Methods* **13**, 573–576 (2016).
85. Qi, R., Ma, A., Ma, Q. & Zou, Q. Clustering and classification methods for single-cell RNA-sequencing data. *Brief. Bioinform.* <https://doi.org/10.1093/bib/bbz062> (2019).
86. Shalek, A. K. & Benson, M. Single-cell analyses to tailor treatments. *Sci. Transl. Med.* **9**, eaan4730 (2017).
87. Zhu, S., Qing, T., Zheng, Y., Jin, L. & Shi, L. Advances in single-cell RNA sequencing and its applications in cancer research. *Oncotarget* **8**, 53763–53779 (2017).
88. Baran-Gale, J., Chandra, T. & Kirschner, K. Experimental design for single-cell RNA sequencing. *Brief. Funct. Genomics* **17**, 233–239 (2018).
89. Suzuki, A. et al. Single-cell analysis of lung adenocarcinoma cell lines reveals diverse expression patterns of individual cells invoked by a molecular target drug treatment. *Genome Biol.* **16**, 66 (2015).

90. Kim, K.-T. et al. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* **17**, 80 (2016).
91. Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
92. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
93. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
94. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).
95. Amodio, M. et al. Exploring single-cell data with deep multitasking neural networks. *bioRxiv* 237065. <https://doi.org/10.1101/237065> (2018).
96. Ding, J., Condon, A. & Shah, S. P. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nat. Commun.* **9**, 2002 (2018).
97. Wang, D. & Gu, J. VASC: dimension reduction and visualization of single cell RNA sequencing data by deep variational autoencoder. *bioRxiv* 199315. <https://doi.org/10.1101/199315> (2017).
98. Risso, D. et al. clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets. *PLoS Comput. Biol.* **14**, e1006378 (2018).
99. Li, H. et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
100. Anchang, B. et al. DRUG-NEM: Optimizing drug combinations using single-cell perturbation response to account for intratumoral heterogeneity. *Proc. Natl Acad. Sci. USA* **115**, E4294–E4303 (2018).
101. Haverty, P. M. et al. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).
102. Hatzis, C. et al. Enhancing reproducibility in cancer drug screening: how do we move forward? *Cancer Res.* **74**, 4016–4023 (2014).
103. Hamilton, W. et al. Inductive representation learning on large graphs. In *Neural Information Processing Systems* 1024–1034 (Curran Associates, Inc., 2017).
104. Kearnes, S., McCloskey, K., Berndl, M., Pande, V. & Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* **30**, 595–608 (2016).
105. Hertel, L., Barth, E., Kaster, T. & Martinetz, T. Deep convolutional neural networks as generic feature extractors. In *2015 International Joint Conference on Neural Networks (IJCNN)*. <https://doi.org/10.1109/ijcnn.2015.7280683> (2015).
106. Zoph, B. & Le, Q. V. Neural architecture search with reinforcement learning. Preprint at <https://arxiv.org/abs/1611.01578> (2016).
107. Li, L. & Talwalkar, A. Random search and reproducibility for neural architecture search. Preprint at <https://arxiv.org/abs/1902.07638> (2019).
108. Rampásek, L., Hidru, D., Smirnov, P., Haibe-Kains, B. & Goldenberg, A. Dr.VAE: Improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btz158> (2019).
109. Zhang, Z. et al. Deep learning in omics: a survey and guideline. *Brief. Funct. Genomics* **18**, 41–57 (2019).
110. Ivanov, A. A., Khuri, F. R. & Fu, H. Targeting protein–protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* **34**, 393–400 (2013).
111. Smirnov, P. et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics* **32**, 1244–1246 (2016).
112. Cokelaer, T. et al. GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btx744> (2017).
113. Rajapakse, V. N., Luna, A., Yamade, M., Loman, L. & Varma, S. Integrative analysis of pharmacogenomics in major cancer cell line databases using CellMinerCDB. *bioRxiv* <https://doi.org/10.1101/292904> (2018).
114. Gupta, S. et al. Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: A step towards personalized medicine. *Sci. Rep.* **6**, 23857 (2016).
115. Mer, A. S. et al. Integrative pharmacogenomics analysis of patient derived xenografts. *Cancer Res.* 471227. <https://doi.org/10.1101/471227> (2019).
116. Lee, J.-K. et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat. Genet.* **50**, 1399–1411 (2018).
117. He, X., Folkman, L. & Borgwardt, K. Kernelized rank learning for personalized drug recommendation. *Bioinformatics* **34**, 2808–2816 (2018).
118. Deshwar, A. G. et al. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16**, 35 (2015).
119. Jiang, Y., Qiu, Y., Minn, A. J. & Zhang, N. R. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **113**, E5528–E5537 (2016).
120. El-Kebir, M., Satas, G., Oesper, L. & Raphael, B. J. Inferring the mutational history of a tumor using multi-state perfect phylogeny mixtures. *Cell Syst.* **3**, 43–53 (2016).
121. Satas, G. & Raphael, B. J. Tumor phylogeny inference using tree-constrained importance sampling. *Bioinformatics* **33**, i152–i160 (2017).
122. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
123. Miller, C. A. et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput. Biol.* **10**, e1003665 (2014).
124. Oesper, L., Satas, G. & Raphael, B. J. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data. *Bioinformatics* **30**, 3532–3540 (2014).
125. Liu, Y. et al. DCDB 2.0: a major update of the drug combination database. *Database* **2014**, bau124–bau124 (2014).
126. O’Neil, J. et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol. Cancer Ther.* **15**, 1155–1162 (2016).

AUTHOR CONTRIBUTIONS

G.A. wrote section on combination therapies, opportunities and challenges for deep learning, edited manuscript. L.R. wrote section on monotherapy prediction. Z.S. wrote section on single-cell drug combination discovery. P.S. wrote section on resistance to monotherapy. B.H.-K. co-wrote introduction and abstract, edited manuscript. A.G. co-wrote introduction and abstract, edited manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41698-020-0122-1>.

Correspondence and requests for materials should be addressed to B.H.-K. or A.G.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020