

Artificial Intelligence in Cancer Research and Precision Medicine



Bhavneet Bhinder^{1,2}, Coryandar Gilvary³, Neel S. Madhukar³, and Olivier Elemento^{1,2,3}

ABSTRACT

Artificial intelligence (AI) is rapidly reshaping cancer research and personalized clinical care. Availability of high-dimensionality datasets coupled with advances in high-performance computing, as well as innovative deep learning architectures, has led to an explosion of AI use in various aspects of oncology research. These applications range from detection and classification of cancer, to molecular characterization of tumors and their microenvironment, to drug discovery and repurposing, to predicting treatment outcomes for patients. As these advances start penetrating the clinic, we foresee a shifting paradigm in cancer care becoming strongly driven by AI.

Significance: AI has the potential to dramatically affect nearly all aspects of oncology—from enhancing diagnosis to personalizing treatment and discovering novel anticancer drugs. Here, we review the recent enormous progress in the application of AI to oncology, highlight limitations and pitfalls, and chart a path for adoption of AI in the cancer clinic.

INTRODUCTION

The term “artificial intelligence” (AI) was first coined for the Dartmouth Summer Workshop in 1956, where it was broadly referred to as “thinking machines.” In simple terms, AI can be defined as the ability of a machine to learn and recognize patterns and relationships from enough representative examples and to use this information effectively for decision-making on unseen data. AI is a vast term that encompasses (and is sometime used synonymously with) machine learning and deep learning. In broad terms, machine learning is a subfield of AI, and deep learning is the subset of machine learning that focuses on deep artificial neural networks (i.e., artificial neural networks with multiple fully connected hidden layers; Fig. 1). In recent years, deep learning has gained enormous traction due to its unprecedented success in computer vision tasks such as face recognition and image classification, among others (1). This property of deep learning extended its applicability to various aspects of cancer research and medicine, such as automatically and accurately detecting cancer from images of stained tumor slides or radiology images, thereby holding the potential to unburden pathologists and radiologists from routine and repetitive tasks.

¹Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, New York. ²Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York. ³OneThree Biotech, New York, New York.

Corresponding Author: Olivier Elemento, Weill Cornell Medicine, Box 140, 1305 York Avenue, New York, NY 10065. Phone: 646-962-5726; E-mail: ole2001@med.cornell.edu

Cancer Discov 2021;11:900–15

doi: 10.1158/2159-8290.CD-21-0090

©2021 American Association for Cancer Research.

Convolutional Neural Networks: Workhorse for Image Classification

Convolutional neural networks (CNN) have been the most popular deep learning architectures used for image classification in cancer (Fig. 1). CNNs apply a series of nonlinear transformations to structured data (such as raw pixels of an image) to learn relevant features automatically, unlike conventional machine learning models that frequently require manual feature curation. On the flip side, it is difficult to tell what features are learned by the CNNs, making them what many have referred to as a “black box.” One consequence is that images used for CNNs should be carefully preprocessed to reduce the risk that the model learns from image artifacts. There are two major approaches for CNN models; one is transfer learning that uses images from a large collection of natural objects (such as in ImageNet) to train the initial layers of a model (where the model learns to identify general features such as shapes and edges) and then uses the disease-specific data to fine-tune the training parameters in the last layers. The second variation of CNN is based on an autoencoder where the model learns background features from a subset of representative images and encodes a compressed representation of the basic features later used to initialize the CNN. In the CAMELYON16 Challenge—a crowdsourced competition to identify and classify lymph node metastasis in patients with breast cancer from whole slide images (WSI) of hematoxylin and eosin (H&E)-stained tumors—25 of the 32 submitted algorithms were CNNs, and the top 5 classification models were exclusively based on transfer learning, which were GoogLeNet, ResNet, VGG-16 (2). Khosravi and colleagues trained and tested several state-of-the-art deep learning models to classify WSI from H&E-stained tumor tissues of The Cancer Genome Atlas (TCGA) cohort and reported on the

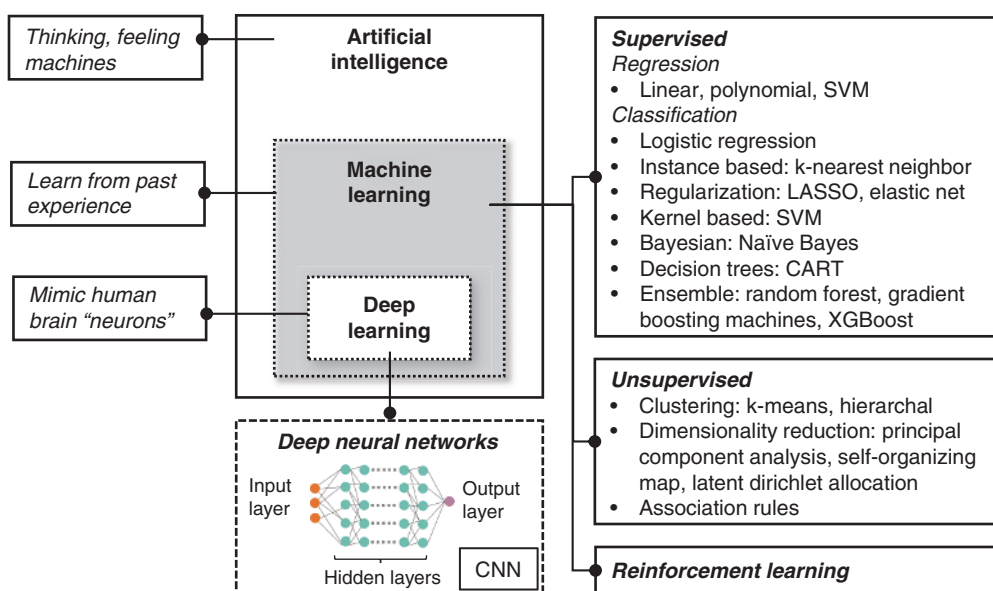


Figure 1. Relationship between AI, machine learning, and deep learning, with commonly used algorithms as examples. CART, Classification and Regression Trees; LASSO, Least Absolute Shrinkage and Selection Operator; SVM, support vector machines.

relative performance of these methods, noting that transfer learning-based inception architectures (GoogLeNet V1 and V3) had an overall best performance for tumor-normal tissue and cancer subtype classification tasks (3).

Generating Predictive Models from Other Large Datasets

In the past decade, several national and international initiatives have resulted in the generation of large cancer datasets. These datasets are obtained from profiling tumor samples using diverse high-throughput platforms and technologies. They are frequently used to build predictive models that inform research and may eventually inform clinical decisions (Fig. 2A). TCGA is by far the most comprehensive publicly available compilation of tumor profiles and includes a large number of data types spanning genomics, epigenomics, proteomics, histopathology, and radiology images (4). Other efforts such as The Pan-Cancer Analysis of Whole Genomes (PCAWG), METABRIC, and GENIE have also compiled large numbers of cancer genomic profiles and made these data publicly available. Profiling technologies have evolved over time. For example, genomic DNA profiling has expanded from targeted panels to whole exomes to whole genomes. Gene expression profiling has evolved from genome-wide microarrays to RNA sequencing (RNA-seq) and then to more granular single-cell RNA-seq (scRNA-seq). Other mature technologies have led to the production of a wide-ranging array of datasets, including DNA methylation profiles, large-scale proteomics studies, perturbation studies including cell viability or cytotoxicity assays using small molecules, RNAi or CRISPR screens, protein-protein interaction networks, and more. The sheer breadth and diversity of datasets that are available publicly or can be generated in minimal time presents a unique opportunity to integrate various data types. Many groups have shown the benefits of such integra-

tion. For example, training predictive models on multiple integrated rather than singular data sources has been shown, for instance by Cheerla and Gevaert, to improve prediction of overall survival in patients across cancers (5). Madhukar and colleagues used such an integrative approach to predict the targets and mechanisms of action of small anticancer molecules and demonstrated clearly that integrating multiple data types improves prediction accuracy (6).

Data Quality and Model Selection Are Key

The basic strategy for machine learning workflows is fairly standard (Fig. 2B). Data collection and cleaning are the first and key components of any workflow, as a model is as good as the data it is trained on. To ensure high quality of the collected data, it needs to be inspected and corrected for possible noise in both non-image (such as inaccurate data entries and missing values) and image (such as high-intensity pixels from artifacts and uneven illumination) data types. The data also need to be reviewed for possible biases that can lead to underfitting the model or high variance that can lead to overfitting the model. A model overfits the data when it learns from artifacts or noise in the data rather than the true signal. The consequence of overfitting is that a model may generalize poorly to unseen data with different biases. Strategies such as cross-validation, increasing the training set size, manually curating predictive features, and using ensemble approaches have been recommended to diminish risks of overfitting.

Another key step of machine learning workflows is to select and fine-tune an optimal model based on its performance. The performance of a machine learning model is commonly measured using the area under the receiver operating characteristics curve (AUROC; or simply AUC), which quantifies the trade-off between sensitivity and specificity. A good classifier should achieve both high sensitivity and high specificity, but emphasis on either of them may be important for some applications. In

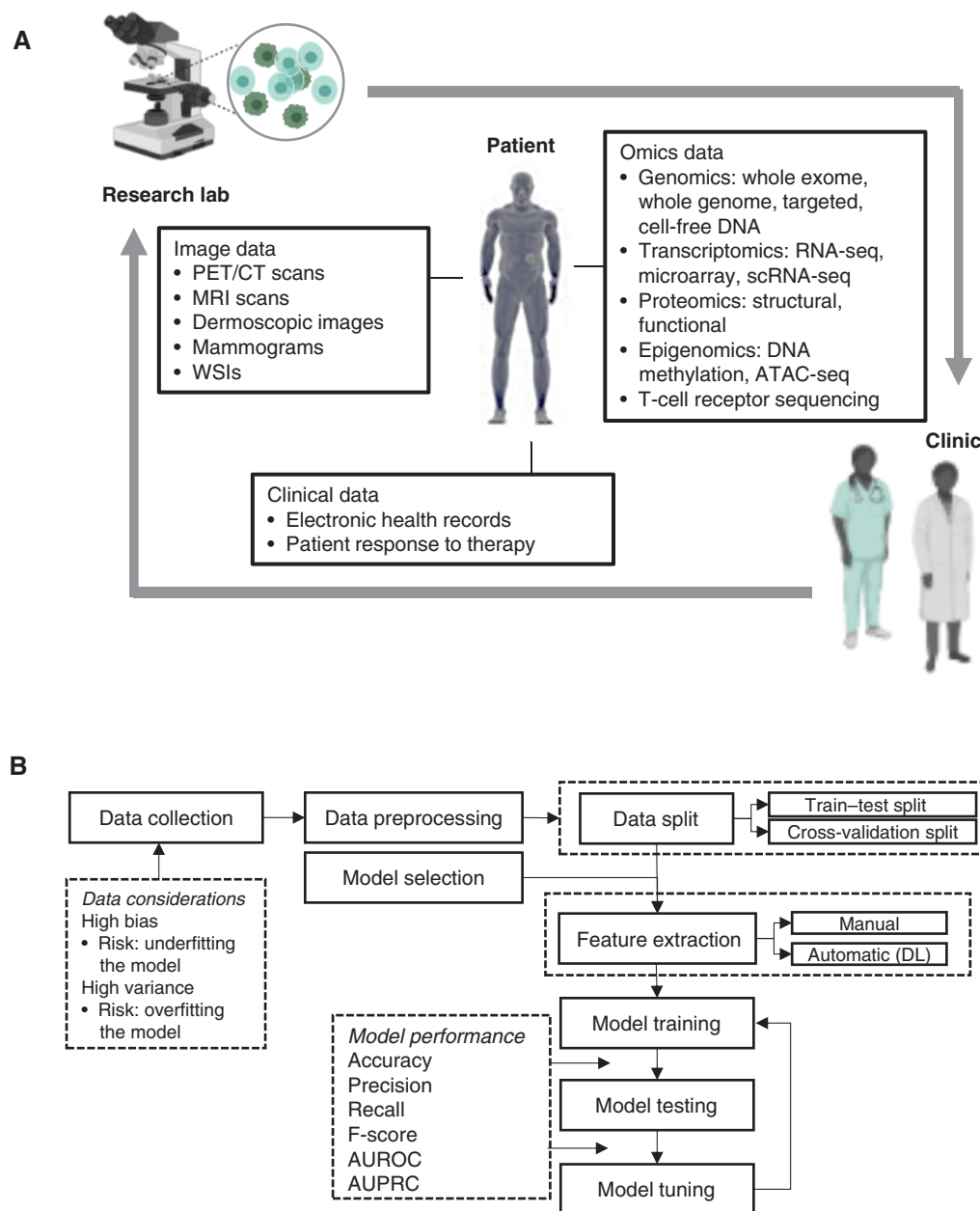


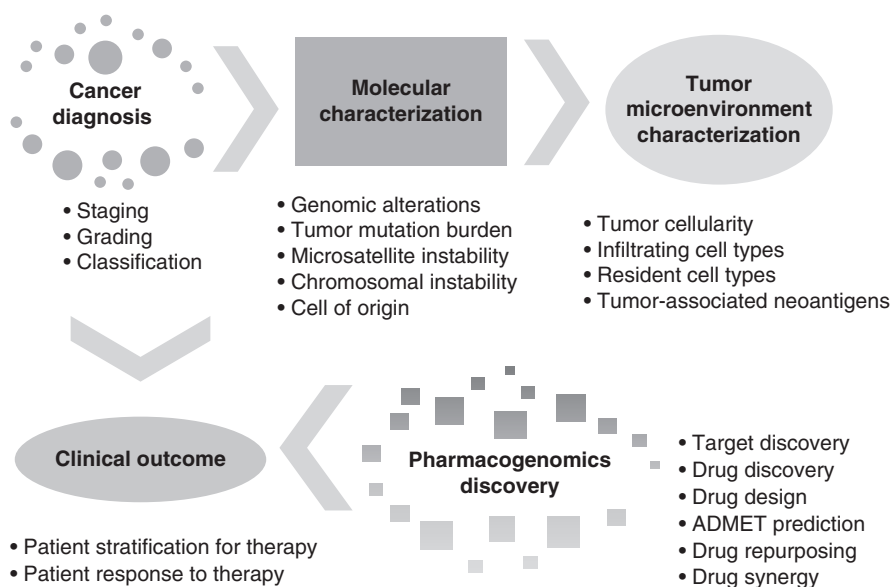
Figure 2. Overview of machine learning basics. **A**, Patient-derived datasets used for AI models with an aim to accelerate bench-to-bedside cancer care. **B**, Basics steps in machine learning classifier workflows. AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristics curve; DL, deep learning.

general, an AUC of >0.80 is considered good, but whether this threshold is also clinically acceptable may vary depending on the clinical use. Even if widely used, there are pitfalls in relying blindly on AUC as a performance metric. For example, the AUC assesses model performance in a population but does not provide confidence in individual calls. For datasets that have a class imbalance such that the positive class (class of interest) examples are much less than the negative class examples and the focus of the model is to accurately detect the positive class, area under the precision-recall curve (AUPRC) is a preferred alternative to AUC. After training and testing a model on a given cohort (usually split into training and test sets), it is equally important to also

validate the model on external independent datasets to ensure that the model is stable and generalizes well. AI model development is not a static process; the model needs to be tested from time to time as newer updated datasets become available. Routine maintenance is frequently required to ensure that model performance does not degrade due to concept drift, that is, when the relationship between the input and output variables changes over time in unforeseen ways.

In this review, we sought to survey a broad spectrum of publications and studies that together capture the breadth and versatility of AI applied to oncology. We sought to describe models that range from those with prospective utilization in

Figure 3. Applications of AI in cancer research and precision medicine. ADMET, Absorption, Distribution, Metabolism, Excretion, and Toxicity.



the clinic to models that drive research and discovery (Fig. 3). This review not only places special emphasis on deep learning as a technique for making machine learning models, but also covers use cases where traditional machine learning techniques have been used very effectively. Finally, we highlight the limitations and challenges that pave the path toward integrating AI models in clinic.

EARLY DETECTION, DIAGNOSIS, AND STAGING OF CANCER

Timing of cancer detection, accuracy of cancer diagnosis, and staging are key determinants of tumor aggressiveness and affect clinical decision-making and outcomes. In just a few years, AI has made significant contributions to this critical area of oncology, sometimes with performance comparable to that of human experts and with an added advantage of scalability and automation.

Making Cancer Diagnoses More Accurate

Deep learning-based models that accurately diagnose cancer and identify cancer subtypes directly from histopathologic and other medical images have been reported extensively. Deep neural networks (DNN) are powerful algorithms that can, with appropriate computing power, be applied to large images such as H&E-stained WSIs of tissue derived from biopsies or surgical resections. These model architectures have indeed excelled at classification of images such as determining whether a digitized stained slide contains cancer cells or not (2, 3, 7–13). While attaining highest prediction accuracies for distinguishing tumor from healthy cells (AUCs > 0.99), DNNs are used for more challenging classification tasks as well, such as distinguishing between closely related cancer subtypes (such as adenocarcinoma vs. adenoma in gastric and colon cancers and adenocarcinoma vs. squamous cell carcinoma in lung tumors) and detecting benign versus malignant tissue. As an example, Coudray and colleagues developed and applied DeepPATH, an Inception-v3 architecture-based model, to concurrently classify

WSI for the TCGA lung cancer cohort into any of the three classes—normal, lung adenocarcinoma, and lung squamous cell carcinoma—with a reported AUC of 0.97 (11).

The success of DNNs is not confined to histopathology images but extends to other medical images acquired through noninvasive techniques such as CT scans, MRI, and mammograms, and even to photographs of suspicious lesions. For example, Esteva and colleagues trained a DNN (Inception-V3 architecture) on skin lesion images labeled for 757 granular skin disease classes (14). Their model, when tested for carcinoma and melanoma classification of photographic and dermoscopic images of skin lesions, outperformed (AUC, 0.91–0.94) the average accuracy attained by 21 board-certified dermatologists. Importantly, their model was robust to variabilities inherent to digital photographs (due to different camera angles, uneven exposures, and so on), hence making the applicability of this model highly generic (14). In radiology, Anthimopoulos and colleagues showed that CT scans of patients with lung disease can be used to build DNNs that classify textural patterns in the lung (such as ground glass opacity and micronodules) with an average accuracy of 0.85 (15). Similarly, Jiang and colleagues used CT scans to develop DNN that predict occult peritoneal metastasis in gastric cancers with an improved AUC (0.92–0.94) compared with that achieved from clinical and pathologic features (AUC, 0.51–0.63; ref. 16). In another work, Wang and colleagues used MRI images from 172 patients with prostate cancer to train and test a DNN (developed using Caffe deep learning framework by Berkeley AI Research) that could distinguish prostate cancer from benign prostate conditions (such as prostate gland enlargement) with a reported AUC of 0.84 (17). In a retrospective study with biopsy-confirmed diagnosis and longitudinal follow-ups, McKinney and colleagues published an ensemble approach with three independent deep learning models that predict cancer risk score directly from the mammograms of approximately 29,000 women (AUC, 0.75–0.88; ref. 18). The group also reported an improvement in absolute specificity (1.2%–5.7%)

and sensitivity (2.7%–9.4%) of cancer detection from mammograms compared with an average radiologist. All in all, such models, if their performance is confirmed in prospective studies, may play an important role in early detection and classification of cancers, especially because their performance is comparable to, if not better than, experts in field. Outside hospital settings, AI-aided smartphone apps have also started to be adopted, potentially bringing early detection of cancerous lesions directly to a user's handheld device (19, 20). However convenient and promising, the diagnostic accuracy of such smart phone applications still remains to be clinically validated. Of particular concern are cases predicted as false negatives, as they may delay patients from procuring timely medical attention (19).

Cancer Staging and Grading

Cancer staging and grading, that is, determining how aggressive and advanced the cancer is, is another important component of the diagnostic process. Staging can indeed affect treatment choices, such as deciding between watchful waiting and aggressive treatment involving radiotherapy, surgery, and chemotherapy. In prostate cancer, staging is achieved using the Gleason score, a combination of two scores measuring prevalence of tumor cells in two distinct locations on a slide. DNNs have shown promising initial results in predicting Gleason scores from histopathology images of prostate tumors (21, 22). Nagpal and colleagues used WSI for H&E-stained prostatectomy specimens to train and test a DNN (Inception-V3) and k-nearest-neighbor classifier-based model to predict Gleason scores (21). The group reported an improved prediction accuracy of Gleason scores estimated from their model (0.70) compared with those determined by a panel of 29 independent pathologists (0.61). Cancer staging can also be done from radiology images: Zhou and colleagues developed a deep learning approach (based on SENet and DenseNet) to predict grade (low versus high) from the MRI images of patients with liver cancer and reported an AUC of 0.83 (22). Overall, these studies indicate promising application of AI to cancer staging, with reported performance on par with trained experts despite modest AUC.

Increasingly, nonimaging data such as genomic profiles are also being used for diagnosis and staging. Data obtained from next-generation sequencing (NGS)—such as whole-exome, whole-genome, and targeted panels; transcription profiles from microarray, RNA-seq, and miRNAs; and methylation profiles—can be used to diagnose cancer and classify tumors into subtypes. Because the data provided by these platforms are highly multidimensional (tens of thousands of genes can be assessed simultaneously), their use for cancer classification requires statistical methods or machine learning (23–25). The use of machine learning for cancer diagnosis and staging from molecular data has in fact been around since the early 2000s, when machine learning approaches such as clustering, support vector machine, and artificial neural networks were applied to microarray-based expression profiles for cancer classification and subtype detection (26). Over the years, omics technologies have advanced and so have the innovations in the machine learning algorithms. Capper and colleagues demonstrated that a random forest classifier trained exclusively on tumor DNA methylation profiles can

significantly improve the prediction accuracies for the hard-to-diagnose subclasses of the central nervous system cancers (AUC, 0.99; ref. 27). Their subclass predictions for 139 cases did not match pathologists' diagnosis, but follow-up of those select cases revealed that approximately 93% of those mismatched cases were in fact accurately predicted by the model (27). Moving into deep learning methods, Sun and colleagues built and applied DNN to genomic point mutations to classify tissues into either of the 12 TCGA cancer types or healthy tissues obtained from the 1000 Genomes Project (28). The classifier, trained on the most frequent cancer-specific point mutations obtained from whole-exome sequencing profiles, was able to distinguish between healthy and tumor tissue with high accuracy (AUC, 0.94), but did not perform as well in a multiclass classification task to distinguish all of the 12 cancer types at the same time (AUC, 0.70). This work highlighted that accurate cancer classification using mutation data is challenging, possibly because of intratumor heterogeneity and low tumor purity (making mutation detection challenging), together with the presence of shared mutations across different cancer types. Nonetheless, the work also shows that similar models that use genomic information to assess cancer can be applied to genomic profiles obtained from other sources such as cell-free DNA (cfDNA).

On the Road to Early Cancer Detection

AI is gradually paving its path toward early detection of cancer from emerging minimally invasive techniques as well, such as liquid biopsies for circulating tumor DNA (ctDNA) or cfDNA. Liquid biopsies, obtained via minimally invasive techniques such as a simple blood test, in theory allow for early detection of cancer, monitoring risk of relapse over time, and guiding treatment options. As an example, microsatellite instability (MSI) status can be predicted from ctDNA in patients with endometrial cancer in order to inform immunotherapy-based treatment (29). Chabon and colleagues developed a machine learning–based approach, Lung-CLiP (cancer likelihood in plasma), that predicts the likelihood of ctDNA in blood drawn from patients with lung cancer (30). The method first estimates the probability that a cfDNA mutation is associated with the tumor (using an elastic net model and features that include cfDNA fragment size) and then integrates outputs of this model together with copy-number scores in an ensemble classifier with five distinct algorithms to predict the presence of ctDNA in a blood sample. The method showed modest predicative performance (AUC, 0.69–0.98), with performance depending on cancer stage, and a tradeoff between specificity and sensitivity for the predictions. In another promising work, Mouliere and colleagues reported a random forest-based classifier trained on features derived from the cfDNA fragment sizes that predicts the presence of ctDNA in blood across multiple cancer types at a high accuracy (AUC, 0.91–0.99; ref. 31). As a complete end-to-end blood test for cancer, Cohen and colleagues developed, for eight distinct cancer types, CancerSEEK, which not only detects early cancer but also predicts any of the eight cancer types directly from the ctDNA (32). Samples are first classified as cancer-positive by a logistic regression model applied to mutations in 16 genes and expression levels in 8 plasma proteins. The cancer type is then predicted using a random

forest classifier (accuracies range from 39% to 84% depending on cancer type; ref. 32). This work is particularly important because five of the eight cancer types covered in this test have no early screening tests currently available. Taken together, the initial progression of AI in the early cancer detection area is notable but has so far been limited to traditional machine learning algorithms. As data acquisition from liquid biopsies expands, we anticipate that more advanced deep learning architectures will eliminate the need for manual selection and curation of most relevant discriminatory features. We also anticipate further use of multimodal approaches (such as CancerSEEK) that combine several data types, e.g., liquid biopsy and imaging, to enhance early detection and monitor disease risk over time.

DETECTING CANCER MUTATIONS USING MACHINE LEARNING

The ubiquitous availability of NGS has made it possible for thousands of cancer laboratories to routinely sequence cancer genes, exomes, and genomes. Identifying genetic variants and mutations in NGS data can be done using a variety of computational tools, but frequently fails in certain scenarios, such as low coverage or complex, repeat-rich regions of the genome. Several groups have explored the idea to recast mutation detection as a machine learning problem (33, 34). As an example, DeepVariant, a DNN (Inception-V2 architecture)-based method, was developed to detect variants from aligned NGS reads by first producing read pileup images for candidate variants (thereby making it an image classification task) and then predicting the probabilities of their genotype likelihood states (homozygous reference, heterozygous variant, or homozygous variant; ref. 33). This method won an award at the second precisionFDA Truth Challenge (2016) for best performance in SNP detection.

Making the Most of Mutations

Another area of interest for AI is the detection of certain key mutations directly from histopathology images, especially clinically actionable mutations that serve as response biomarkers for targeted therapies (such as activating mutations in *EGFR*). This would offer a cost-effective and faster alternative to mutation detection from NGS, as it would leverage ubiquitously available image data, from both pathology and radiology. DeepPATH, besides classifying subtypes of TCGA lung cancer, was also able to identify six key mutations in lung cancer, that is, *STK11*, *EGFR*, *FAT1*, *SETBP1*, *KRAS*, and *TP53* (as reported from whole exomes) directly from the WSI of 59 patients at AUCs that ranged between 0.73 and 0.85 (11). The results were promising, but understanding of what features are being learned by the DNN models to determine mutation status for each slide still remains wanting. The group also tested their model to detect *EGFR* mutations in an independent lung cancer cohort and obtained a lower AUC of 0.687. They attributed this lower AUC to differences in sequencing platform and tissue preservation techniques between the independent cohort and the TCGA cohorts (on which their model was trained and validated). Following on this work, other groups have also applied AI approaches to identify mutations from images. For example, a trans-

fer learning-based DNN approach could determine *EGFR* mutation status directly from preoperative CT scans of 844 patients with lung adenocarcinoma with AUC > 0.81 (35). Determination of *EGFR* mutation status in non-small cell lung cancer tumors was also achieved directly from ¹⁸F-FDG-PET/CT scans using SResCNN model with AUC > 0.81 (36). Driver mutations (e.g., *IDH1*) and MGMT methylation status could be detected in diffuse low-grade gliomas using MRI images for feature extraction followed by XGBoost Model with AUC > 0.70 (37). DNN (Inception-V3) could identify common mutations in liver cancer (*CTNNB1*, *FMN2*, *TP53*, and *ZFX4*) directly from WSI using with AUC > 0.71 (38).

The focus on somatic mutations has expanded from assessing mutations in individual genes to assessing mutational footprints, that is, the number and context of all mutations found within a tumor. MSI status is an example of mutational footprint in tumors that has gained a prominent role as a diagnostic and predictive biomarker for checkpoint immunotherapies (39). As an example, the FDA recently approved Keytruda (pembrolizumab) as a first-line treatment for patients with MSI-high (MSI-H) metastatic colorectal tumors (<https://www.fda.gov/news-events/press-announcements/fda-approves-first-line-immunotherapy-patients-msi-hdmmr-metastatic-colorectal-cancer>). This has spurred the search for fast and cost-effective methods that can easily detect MSI-H tumors. As before, one compelling idea would be to predict MSI status directly from H&E-stained histopathology images, which are readily available and do not require additional tissue; this would provide a cost-effective and time-sensitive alternative to existing methods, for example MSI inference from qPCR, immunohistochemistry (IHC), or NGS. With that goal in mind, Kather and colleagues applied ResNet18 CNN to first detect tumor regions in H&E slides (AUC > 0.99) and then classify them as either MSI or microsatellite-stable. This method was applied to 1,600 TCGA tumors focused on gastric, colorectal, and endometrial cancers (40). Model performances were cancer-dependent, with AUCs ranging around 0.75 to 0.84. Interestingly, analysis of formalin-fixed, paraffin-embedded (FFPE) slides was associated with better prediction accuracy (AUC = 0.84) compared with the snap-frozen slides (AUC, 0.77). Validation in an external colorectal cancer cohort gave a comparable performance in accuracy (AUC, 0.84). Interestingly, their method did not perform as well in a different gastric cancer cohort that had individuals of a different ethnicity (Asian, $n = 185$) than the ones used to train the models (TCGA-STAD is predominantly non-Asian; AUC, 0.69; ref. 40). In more recent work, Yashmita and colleagues trained and tested MSINet, a transfer learning model based on MobileNetV2 architecture, to classify tissue and subsequently classify MSI status in H&E-stained histopathology slides (40× magnification) from a colorectal cancer cohort of 100 primary tumors from Stanford Medical Center (41). The group reported an AUC of 0.93, which is a good improvement over the previously reported ResNet18 model (40, 41). Yashmita and colleagues compared their model with the previously published ResNet18 model in two ways: (i) They retrained the ResNet18 model on their internal cohort ($n = 100$) and applied it to the TCGA-CRC cohort ($n = 479$): here they show that ResNet had an AUC of 0.71, whereas their model MSINet had an AUC of 0.77 (or AUC of 0.83 when

restricted to 40× magnification only), or (2) they retrained their model MSINet on the same training set as used by Kather and colleagues and applied it to their internal dataset ($n = 100$): here they report an AUC of 0.88 for their model versus an AUC of 0.77 for ResNet18 model (41). Both of their comparative strategies showed an improved performance of MSINet compared with the ResNet18 model, and the generally lower AUCs for TCGA cohort may be the result of high heterogeneity in the TCGA datasets, which are gathered from multiple institutions.

Tumor mutation burden (TMB) is another important biomarker of response to checkpoint immunotherapy (42). Normally estimated using NGS and thus at high cost and with high variability across platforms and gene panels (43), its estimation directly from histopathology slides is also becoming an area of active research. As a first attempt to determine TMB directly from WSI, Jain and colleagues reported a deep learning model based on Inception-v3 architecture, Image2TMB, to determine the TMB status (high vs. low) from frozen H&E slides in the lung adenocarcinoma (LUAD) TCGA cohort ($n = 499$; ref. 44). The model was trained and tested at three magnifications (5×, 10×, and 20× magnification), and the TMB status probabilities from the three magnifications were aggregated using a random forest model to predict if the TMB is above or below their predefined TMB (AUC, 0.92). In another work, Wang and colleagues also attempted to classify TMB status from FFPE slides for the gastrointestinal cohorts from the TCGA ($n = 545$; ref. 45). Like Jain and colleagues, this group also relied on TMB calculated from nonsynonymous mutation counts from whole exomes and used the upper tertile as the cutoff to define high TMB. The group compared eight different transfer learning models and reported GoogLeNet as their best model for gastric tumors (AUC, 0.75) and VGG-19 as their best model for colon tumors (AUC, 0.82). Besides histopathology images, CT scans have also been used to predict TMB in non-small cell lung cancers (AUC, 0.81; ref. 46). Related to TMB, researchers are now seeking to predict chromosomal instability, a known driver of cancer evolution, directly from histopathology slides (47).

Determining Tumor Cells of Origin

Clinically, determining the cell of origin of tumors can inform site-specific therapies, which have been reported to be more effective than systemic chemotherapies (48). This is relevant for those tumors where the primary sites are unknown, or for cfDNA obtained from liquid biopsies. Different tumor types have distinct patterns of somatic mutations, and these patterns can be leveraged to identify the tissue of origin for tumors. Conventionally, tissue of origin is determined using approaches that include IHC and gene expression profiling assays, but the accuracy of these methods is estimated to be about 80%, wanting further improvement (49). As an alternative, Jiao and colleagues as a part of the PCAWG Consortium built and applied multiclass DNN-based models to binned mutation counts obtained from whole genomes of approximately 6,000 tumors spanning 28 cancer types including primary and metastatic tumors (50). The basic idea behind the approach is that the regional mutation counts are representative of the chromatin accessibility of the genomic region and therefore may recapitulate the epigenetic state of the cell

of origin. Specifically, they show that the regional distribution of somatic mutations per Mb bins across the genome, the majority of which are passenger mutations, can accurately predict the tissue of origin (overall accuracies 0.83–0.91; accuracies varied highly among tumor types). Interestingly, the presence of driver genes or pathways was not found to be a useful classification feature in this model.

Altogether, the methods discussed in this section highlight the growing potential of AI to detect cancer mutations. Although such methods may not be accurate enough for replacing molecular pathology assessment, they may help shed light on cellular mechanisms associated with mutations and may help screen a large number of patients and tumors for subtypes likely to have specific mutational profiles. This may in turn help in designing clinical trials and identifying groups likely to benefit from specific targeted therapies. We anticipate many more complementary methods to be developed in the future. For example, AI may increasingly be used to help understand the functional impact of mutations, e.g., predict the impact of noncoding mutations on gene expression, epigenetic processes, as well as disease risk (51, 52). In coming years, we also anticipate that the detection of mutations from histopathology images may gain further clinical relevance. For example, it may be possible to predict resistance to therapy, changes in mutation status, and, broadly speaking, tumor evolution directly from the histologic pattern changes in pathology images collected from longitudinal tumor specimens (53–54).

CHARACTERIZING THE TUMOR MICROENVIRONMENT

Despite consistently high predictive performances, many of the AI approaches used in digital pathology can be described as “black box”; that is, AI methods can be taught to discriminate between different types of diseases, but often do not provide an easily interpretable explanation underlying the classification process. This is unlike the process used by trained pathologists, who use well-documented features of images and cell morphology and decades of training to assess tissue. AI has the potential to help automate that process and simplify routine tasks that may be relatively time-consuming for a pathologist, for example estimating the quantity of tumor cells in a tissue or determining the cell of origin for a given specimen from its tissue morphology. Tumor cellularity, that is, the fraction of tumor cells in a specimen, is an important indicator of residual disease (pathologic response) after therapy. On a more practical level, tumor cellularity estimation also helps pathologists select appropriate tissue blocks for further analysis, e.g., genomic sequencing. Traditionally, pathologists inspect stained tissue slides to determine tumor cellularity, an approach that is not just laborious but also highly subjective due to intraobserver and interobserver variability. Tumor cellularity can also be inferred computationally from NGS datasets, but there is limited concordance among available inference methods and heavy dependence on the presence of high numbers of genomic alterations for adequate accuracy (55). To address this task using an AI approach, Akbar and colleagues aimed to quantify tumor cellularity directly from H&E-stained WSI (20× magnification) from

53 patients with breast cancer using DNN (InceptionNet architecture), eliminating the need for nuclei segmentation and classification, and feature extraction (56). The group trained two DNN models, one to distinguish tumor from healthy tissue, and the other to output regression scores (between 0% and 100%) indicative of tumor cellularity. Their predicted scores had a good concordance with the tumor cellularity reported by two independent pathologists (correlation 0.82; ref. 56). Although these initial findings demonstrate the feasibility of quantifying tumor cellularity directly from WSI, the models need to be trained and tested on larger datasets.

Further extending the analysis of tumor purity, AI approaches are being used for the spatial and quantitative assessment of the tumor microenvironment (TME). Tumor cells constantly interact with other cells in their microenvironment, such as immune and stromal cells, and these interactions partly determine how tumors evolve, metastasize, or respond to therapies (57). Characterization of the TME is therefore important to investigate these mechanisms. Such characterization is especially important for understanding tumor-immune cross-talk in the context of checkpoint immunotherapies. Saltz and colleagues demonstrated the feasibility of identifying and quantifying lymphocyte infiltration directly from H&E-stained histopathology slides acquired for 13 TCGA cancer types using a DNN with convolutional autoencoder, where the autoencoder learns a compact representation of basic morphologic features (such as cell nuclei and lymphocytes) from the pathology slides and uses this to initialize the neural networks for training (58). The group trained two DNNs, one to classify tumor-infiltrating lymphocyte (TIL) status of each patch in a given image, and the other to identify regions of necrosis on the slide so as to reduce false positives. The patches were later aggregated and manually inspected by pathologists to refine the model outputs. The fraction of TIL-positive versus TIL-negative patches in a slide was then quantified. Using a subset of pathology-assessed lung tumors patches (LUAD) as gold standard, they reported an AUC of 0.95. In another work, Bejnordi and colleagues trained and tested a DNN (VGG-Net architecture) on histopathology images from breast biopsies of 882 patients to distinguish benign from malignant tissues and classify normal versus tumor-associated stroma with an accuracy of 92% (59). Recently, Fassler and colleagues leveraged histopathology images obtained from multiplex IHC of pancreatic cancer tissue and applied a DNN comprised of an autoencoder (ColorAE) together with a U-Net CNN (60). Cell segmentation and classification performance ranged from 0.40 to 0.84 (expressed as F1 score, an alternative to AUC). In the future, multiplexed imaging platforms (such as Vectra PerkinElmer and imaging mass cytometry) capable of imaging multiple aspects of the TME at rapidly increasing resolution will increasingly be used, together with deeper network architectures (such as GoogLeNet and Inception-V3) and more powerful graphics processing units. These technologies will allow researchers to study in detail complex cell-cell interactions within the TME.

Besides using histopathology slides to determine the composition of the TME, DNNs have also been used to deconvolve bulk RNA-seq or microarray profiles into repertoires of resident or infiltrating cell types, based on data obtained

from scRNA-seq profiles. These methods, which include ScanDen and DigitalDLSorter (61, 62), are powerful but of limited use, because currently single-cell profiles from only a small subset of tissue types are publicly available. Nonetheless, these gaps are being addressed, from using higher-throughput solutions for scRNA-seq (such as 10× Chromium) to coordinating global initiatives such as The Human Cell Atlas that aim to comprehensively profile every cell type of the human body (63).

Studies that focus on improved quantifications of individual cell types in the TME, especially the immune cells as described above, are gaining interest mainly due to the success of checkpoint immunotherapy in clinic. Indeed, the TME plays a major role in mounting an antitumor immune response, especially when immune cells are already primed by immunogenic tumor-associated neoantigens. Neoantigens are mutated peptides that arise from tumor-specific mutational events (nonsynonymous mutations, truncating mutations, novel gene fusions, and alternate splicing) and are recognized as nonself by the patient's immune system. Neoantigens are studied extensively for their role in driving exceptional response in patients treated with checkpoint immunotherapies and their potential use in adoptive T-cell therapy and personalized peptide vaccines (57). As standard practice, mutations detected from exome or genome sequencing are collected and translated *in silico* into corresponding mutated peptides. Neoantigens are then inferred from these mutated peptides by predicting their binding affinities to the patient's MHC class I alleles. One of the earliest and state-of-the-art neoantigen prediction tools, NetMHC, is based on artificial neural networks. Among the other existing MHC-peptide binding prediction tools, the majority are still based on artificial neural networks (MHCflurry and EDGE), whereas some of the newer approaches have expanded to other models, such as random forests (ForestMHC), more advanced AI algorithms such as natural language processing (NLP; HLA-CNN), or CNN (ConvMHC and DeepSeqPan), sometimes directly trained from raw data in immunoprecipitation assays or mass spectrometry (MS; refs. 57, 64). As the trend for clinical use of neoantigens shifts toward a more personalized approach, for example with personalized peptide vaccines, Tran and colleagues developed a patient-specific methodology where NLP models are trained on a patient's wild-type immunopeptidome and then applied to the patient's mutated immunopeptidome in order to predict *de novo* peptide sequences of likely neoantigens (65). The model needs broader validation because it was trained and tested on only five patients with melanoma. Nonetheless, the group presented a highly personalized and exciting approach to predict HLA-bound neoantigen sequences directly from a patient's MS data without dependence on NGS for mutations or MHC allele predictions. Whether these predicted neoantigens, from this and other methods above, are truly immunogenic still needs to be experimentally tested.

DISCOVERY OF THERAPEUTIC TARGETS AND DRUGS

Drug discovery and development is often associated with elevated costs and time burden. Affordable access to various

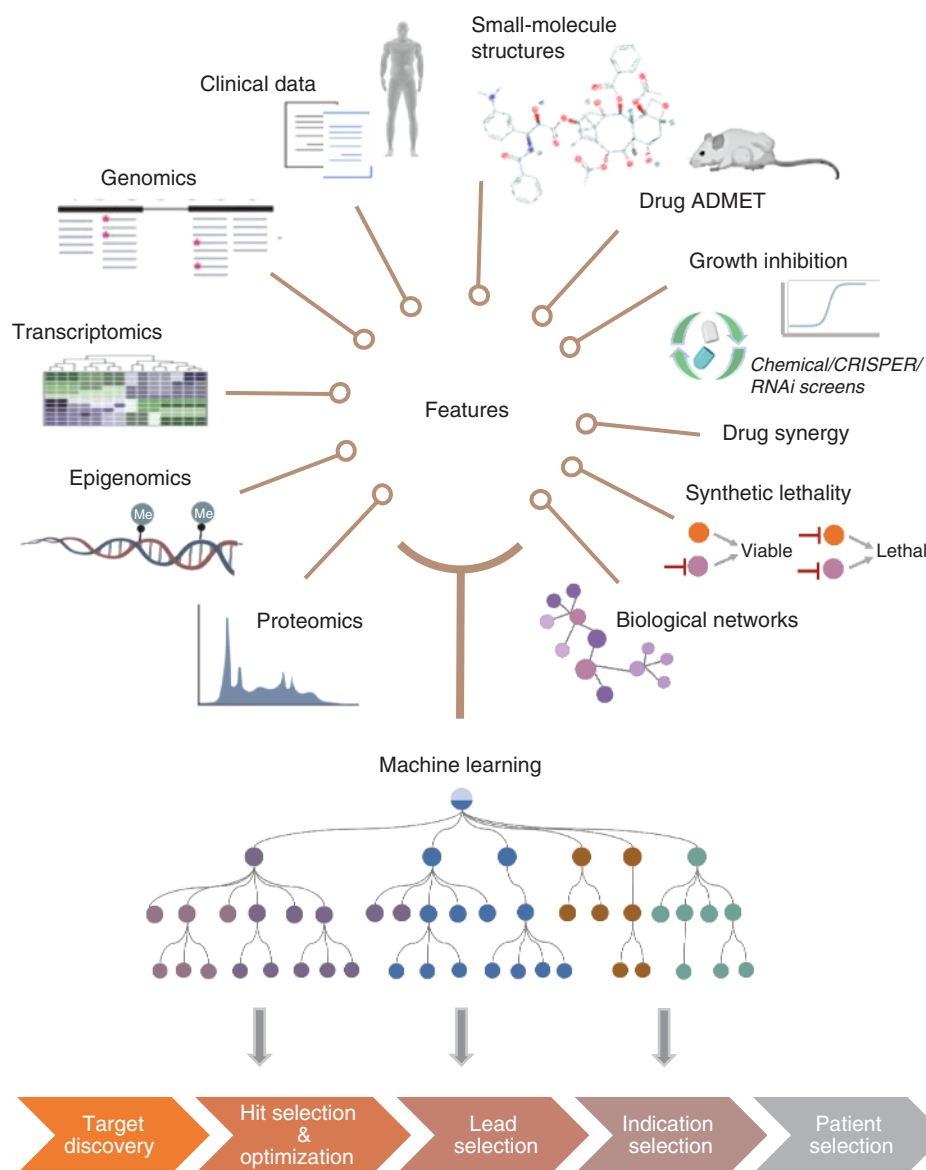


Figure 4. Integrating datasets from multiple diverse sources, ranging from sequencing to structure to screening datasets, can increase the relevant feature space for AI models that enable end-to-end drug discovery.

NGS and imaging technologies together with a growing availability of large cancer datasets (public or private) has led to an exploding interest in leveraging AI to make this process more efficient. This includes developing models that integrate diverse datasets to address each component in the drug discovery spectrum (Fig. 4). As an example, Tong and colleagues integrated clinical data with gene expression profiles and protein–protein interaction networks to derive features that could predict candidate drug targets in liver cancer using one-class support vector machine (AUC, 0.88; ref. 66). In a breast cancer-specific deep learning-based classification approach, López-Cortés and colleagues integrated numerous cancer databases such as PharmGKB, Cancer Genome Interpreter, and TCGA, among others, to predict proteins associated with breast cancer pathogenesis, and

reported several viable candidates to pursue as biomarkers or drug targets (4, 67–70). The DepMap Consortium has made hundreds of loss-of-function screen datasets available to researchers that enable implementation of diverse AI strategies (71). For example, the ECLIPSE machine learning approach predicts cancer-specific drug targets based on the DepMap data by leveraging both gene-specific and cell line-specific data (72). Similarly, Chen and colleagues examined a wide breadth of molecular features from DepMap data and found that proteomics data (specifically, reverse-phase protein array data) are highly predictive of cancer cell line dependencies (73). This finding underscores the versatility of AI to not only predict therapeutic targets, but also assess the type of experimental data most relevant to a predictive model.

Drug Design

AI has also been applied to design drug structures *in silico* with desired physiochemical properties and target specificities. Traditional AI techniques have focused on binary classification and have difficulty modeling complicated objectives, such as generating new molecules *in silico*. Reinforcement learning, a growing subset of AI that is ideal for problems with complex objectives and allows for interactive feedback, has been heavily used within *in silico* molecule generation (74–76). Olivecrona and colleagues demonstrated how their recurrent neural network approach, tuned using policy-based reinforcement learning, was capable of generating analogues to celecoxib and compounds without the element sulfur (74). You and colleagues introduced a graph convolutional network approach that used reinforcement learning to generate novel molecules, showing high accuracy when optimized for a specific property or when creating analogues with certain properties (76). The use of graph convolutional networks has especially affected and improved molecule generation because it can better model chemical molecules and does not require computational conversion of molecules to their two-dimensional representations. Besides, generative adversarial networks (GAN), a combination of two networks—the generator and the discriminator—to build a stronger generator model, has also been commonly applied for molecule generation tasks (77, 78). MolGAN, a method for generating molecules with specific properties, used both GAN and reinforcement learning architecture and achieved high performance for various properties, including drug likeliness, synthesizability, and solubility (62%, 95%, and 89%, respectively; ref. 78). Although neural network-based models dominate this area of molecule generation, nonneural network-based models have been successful in the area of predicting drug properties (79–81). Gayvert and colleagues published a random forest model that used distinct preclinical data types to predict drug toxicity and adverse events (79). Shen and colleagues trained a support vector machine model to predict various absorption, distribution, metabolism, and excretion properties of a drug and validated their approach by accurately predicting both the blood–brain barrier permeability and the human intestinal absorption (81).

Drug Repurposing

Drug repurposing—finding new therapeutic use for an “old” drug beyond its existing medical indication—offers a speedy, safe, and economic alternative to conventional drug discovery. New initiatives such as Library of Integrated Network-Based Cellular Signatures (LINCS) have released rich transcriptional datasets (such as gene perturbation profiles) that can be leveraged by AI to accelerate drug repurposing efforts (82). LINCS datasets, along with others, have been used to identify repurposing candidates from drugs that can reverse the expression profiles of cancer-specific gene signatures (obtained by comparing expression of cancer cells with normal cells; refs. 83–85). DNNs trained on drug-perturbed transcriptional profiles from LINCS have also been used to predict the therapeutic use category for drugs (e.g., vasodilator, antineoplastic) and to prioritize repurposing candidates by their chemical structural similarity with approved cancer drugs (86). In addition to

transcription profiles, publicly available datasets obtained from cell viability assays (that measure the amount of metabolically active cells after treatment with a specific molecule) have also been used to train AI models [Genomics of Drug Sensitivity in Cancer (GDSC), PRISM, NCI-60, etc.; refs. 87–89]. CDRScan, an ensemble of five CNN-based models trained on cell viability datasets from GDSC and the COSMIC cell line project (CCLP), predicts which drug from the GDSC would be most effective for a patient based on their individual somatic mutation profile (90). Besides cancer-specific repurposing efforts, there are numerous other disease-agnostic approaches for drug repurposing that can be extended to cancer (91–93). DeepDR, a variational autoencoder-based DNN model, predicts novel drug–disease connections, based on known clinical annotations and chemical structures of drugs (92). Similarly, Gottlieb and colleagues created PREDICT, a computational pipeline, to predict novel indications for drugs based on integrating both drug and disease similarities (93). PREDICT identified numerous novel indications for known therapies, including the use of progesterone for a rare form of renal cell carcinoma, an association that has support in the literature. The identification of repurposing candidates is an active area of research in AI and by now has led to many promising models and repurposing predictions.

PATIENT PROGNOSIS AND RESPONSE TO THERAPY

The ability to prospectively identify patients best matched for a given therapy can help reduce risks of poor clinical outcomes and also help reduce high costs of treatment, which can average up to \$150,000 a year. This is especially relevant for checkpoint inhibitor immunotherapies, where favorable response rates are low overall (approximately 20%), but certain patients show exceptional, long-term clinical benefit. The use of AI in this area has been limited due to insufficient data availability but is now gradually expanding. Liu and colleagues reported a logistic regression–based classifier trained on treatment-naïve genomic and transcriptomic profiles and clinical features to predict resistance to PD-1 inhibitors in patients with advanced melanoma (AUC, 0.73–0.83; ref. 94). Litchfield and colleagues compiled the largest cohort thus far of matched genomic and transcriptomic profiles from published checkpoint inhibitor studies and used this dataset to train and test a XGBoost-based cancer-specific classifier for prediction of response to checkpoint immunotherapies (AUC, 0.66–0.86; ref. 95). Johannet and colleagues reported a more advanced AI approach using CNNs trained and tested on treatment-naïve histopathology slides together with patients’ clinical characteristics to predict responses to checkpoint immunotherapy in patients with advanced melanoma (AUC, 0.80; ref. 96).

Aside from immunotherapies, models that predict patient responses to other cancer treatments from omics or image data have also been widely reported. Sun and colleagues applied DNNs to features extracted from gene expression, copy-number alteration, and clinical profiles of patients with breast cancer (from METABRIC and TCGA) to predict patient prognosis after treatment with varied indications (AUC > 0.80; ref. 97). Similar omics-based approaches that

use DNN have been shown to predict patient survival from gene expression and pathway profiles in brain cancer and to predict patient survival from gene expression, miRNA expression, and methylation profiles in liver cancer (98, 99). Regarding image-based models, Korfiatis and colleagues applied a DNN (ResNet architecture) model to preoperative MRI scans for brain tumors to predict the methylation status of the *MGMT* gene (which is an established biomarker for patient prognosis after surgery or therapy); the ResNet50 model showed good predictive performance in validation sets (F1 score = 0.95–0.97; ref. 100). CNNs have also been applied to preoperative or pretreatment CT scans to predict disease-free survival in patients with lung cancer (101, 102). In another work, Mobadersany and colleagues trained a CNN with a final layer of Cox regression model to predict patient risk directly from histopathology slides in brain tumors (median c index = 0.75), and the performance of their model improved further after inclusion of genomic markers (isocitrate dehydrogenase mutation status and 1p/19q codeletion) in the CNNs (median c index = 0.801; ref. 103). Similarly, Bychkov and colleagues applied CNNs to predict survival from tissue patch images of H&E-stained histopathology slides in patients with colorectal cancer who underwent surgery (HR, 2.3 for predicted patient stratification; AUC, 0.69). Surprisingly, the model performed better than the consensus assessments provided by three human experts (HR, 1.67; AUC, 0.58; ref. 104). Skrede and colleagues applied CNNs (MobileNetV2 architecture) to the H&E-stained WSI of resected tumors to directly predict patient prognosis in response to chemotherapy and/or radiotherapy (or none) in early-stage colorectal cancer (AUC, 0.71); multivariate survival analysis between patient groups stratified based on the model's predictions shows that the patients predicted to have a poor prognosis indeed had poor cancer-specific survival (adjusted HR, 3.04) compared with those with predicted good prognosis (105).

It is also important to identify early on if an ongoing therapy is not effective for a patient, and if the clinician needs to switch or alter the course of treatment in time. In the clinical setting, cancer progression and response to therapy are monitored by manually inspecting pathology or radiology images to quantify tumor shrinkage and to check for appearance of new lesions. This manual assessment can however be challenging especially for checkpoint inhibitor immunotherapies where patterns of disease progression can be atypical (106). To this end, Dercle and colleagues showed the possibility of using machine learning to train models on treatment-specific features to predict response to distinct cancer treatments (107). The group used an ensemble of six machine learning algorithms to predict patient sensitivity (defined as progression-free survival above the population median) to chemotherapy, targeted therapy, and immunotherapy, using quantitative features extracted from longitudinal CT scans of patients with non-small cell lung cancer (AUCs of 0.67, 0.82, and 0.77 respectively; ref. 107). In another work, Choi and colleagues applied CNNs to predict response to neoadjuvant chemotherapy in patients with advanced breast cancer from PET/MRI scans of both treatment-naïve and chemotherapy-treated tumors; the predictive performance of their model (AUC, 0.60–0.98) was reportedly better than certain conventional methods of response prediction (such

as the difference in standardized uptake volume quantified from the serial CT scans before and after treatment; ref. 108). In a more focused time series model, Xu and colleagues used CNN with recurrent neural networks applied to longitudinal CT scans of lung tumors to predict overall survival in patients after chemoradiation (AUC, 0.74; stratified patient HR, 6.16; ref. 109). In addition to monitoring patient responses to therapies, machine learning models such as CURATE.AI now offer additional avenues to adjust drug dosage for single or combination therapies for individual patients in a dynamic manner using patient-specific data points collected over time (110).

PREDICTING DRUG EFFICACY AND SYNERGY

More broadly, machine learning algorithms have been applied to predict drug efficacy based on molecular features. This work has gained importance due to availability of large cancer drug efficacy datasets, obtained from experiments done in cell lines (87, 89, 111, 112). Although cell lines are imperfect models due to genetic drift or cross-contamination (113), they provide AI models with a large quantity of data to learn from. As with all datasets, preprocessing often needs to be performed to minimize potential noise, such as cell line authentication or validation of *in vivo* data (114). In one study, Iorio and colleagues measured the response of 1,001 cancer cell lines to 265 different anticancer compounds (115). Based on those results, they built a series of Elastic Net models to translate genomic features such as mutations and gene expression values into drug efficacy (in the form of IC_{50} values). The models were able to accurately predict efficacy. Owing to both their accuracy and interpretability, random forests are a commonly used method for drug response prediction and have been shown to improve overall accuracy compared with other machine learning approaches (116). Besides traditional machine learning, deep learning is also becoming a widely used choice for drug response prediction. Using data from TCGA and the Cancer Cell Line Encyclopedia, Chiu and colleagues trained a set of three DNNs to predict drug response: one built to encode mutation information, one built to encode expression information, and a drug response predictor network integrating the first two DNNs (117). They found that this method was able to identify both known and novel drug–cancer pairings, and interestingly, they found that expression data contributed more to accurate predictions than mutation information. Also using a DNN, Sakellaropoulos and colleagues trained a model using GDSC cell lines and then applied it to various genomic datasets with clinical response data (118). Using predicted IC_{50} values, they split patients into high-sensitivity and low-sensitivity cohorts and found that their DNN was able to separate patients based on survival under certain treatment regimens.

One of the biggest drawbacks in using deep learning is that most methods suffer from a lack of interpretability into the underlying biological mechanisms that drive the prediction. To address this, Kuenzi and colleagues developed DrugCell, an interpretable deep learning model that uses a “visible neural network” (VNN) to ensure that the underlying neural network hierarchy resembled known biological processes (119). They combined this VNN with an artificial neural network built to model a drug's chemical structure

and found that this combination could correctly predict drug response (spearman $\rho = 0.80$ when comparing predicted efficacy vs. actual efficacy) while also providing insight into the underlying mechanisms of action-driving response. They also showed how this approach could be used to predict synergistic drug combinations and validated their predictions in patient-derived xenograft models with an AUC of 0.75.

Following on DrugCell, some approaches have sought to combine genomic data with other features to predict single-drug or combination efficacy. Cortés-Ciriano and colleagues showed that by combining chemical information with biological information (genomics, transcriptomics, and proteomics) on specific cell lines, they were able to predict efficacy for 17,000+ compounds across the 59 cell lines in the NCI-60 dataset (116). Extending this analysis to drug combinations, a recent DREAM challenge crowdsourced different models to predict drug synergy in a subset of cancer cell types. With more than 80 distinct models submitted, they found that those that integrated genomic features with other information (such as a chemical structure or known biological interactions) tended to produce higher overall accuracies (120). Similarly, Gilvary and colleagues have also reported that, using a multitask suite of models that integrate genomic, target, chemical, and effect-based features, they can retain high predictive accuracy while also deconvoluting the mechanisms that may be contributing to the predicted synergy (121).

CURRENT CHALLENGES AND FUTURE PERSPECTIVES

AI has indisputable potential to enhance the care of patients with cancer and more broadly affect the field of cancer. In the laboratory setting, it has shown performance accuracies high enough to, in theory, transform conventional practices at almost all stages of cancer research and medicine (Fig. 3). After the tremendous success of AI at the bench, the question becomes whether, and then when, AI can become fully integrated in the clinic as a regular practice for doctors and patients with cancer.

AI runs on data; in the clinical setting, data that adequately capture the entire human population are key to developing robust AI models. It is becoming increasingly clear that differences in race and gender together with socioeconomic disparities affect disease risk and recurrence among individuals. In cancer, race-specific variations in occurrence and frequency of genomic aberrations have been reported (122). Work by Bhargava and colleagues has in fact shown that race-specific differences exist even at the level of tissue morphology—and so do differences in disease aggressiveness—between Caucasian and African American men with prostate cancer (123). But existing datasets that are commonly used to train and test AI models in cancer are still inherently biased toward certain racial and ethnic groups. As an example, TCGA, the largest repository of varied cancer datasets, is predominantly composed of white individuals with European ancestry (122). Other biases exist within the commonly used large datasets. For example, the TCGA cohorts are mainly comprised of primary tumors with a very limited availability of metastatic tumors. Cell lines, which are the workhorses of preclinical drug development and frequently populate large genomic datasets, do not capture

the real-world patient profiles accurately, as they are prone to issues such as genetic drifts (which is divergence in the genome due to multiple cell line passages). As patient-derived organoids become more readily available, cell line-based datasets will be complemented with experimental data obtained from these organoids, which are genetically more stable (124). Aside from data biases, there are also gaps between ease of data acquisition from various platforms versus ease of data access by external institutions for independent use, especially for private or controlled-access datasets. As clinical studies and associated datasets of the future continually evolve to become more inclusive, harmonized, and easily accessible, these data chasms that challenge robust clinical implementation of AI will also be bridged.

In addition to data sharing, code sharing for AI models is another aspect that would ensure that the models are transparent and reproducible and are good candidates for clinical use. For most published studies, authors do validate their models on external datasets, but for their models to be truly translatable and clinically relevant, they should be independently reproducible in the hands of others, just like any other credible scientific finding. This can be made possible by sharing well-documented code for the model together with transparent descriptions of the optimized hyperparameters and hardware specifications. But as Haibe-Kains and colleagues point out, despite multiple available options for code sharing (such as Github) with version-controlled virtual environments (such as Docker), sharing well-annotated code for complex models is still not universally adopted (125). Thankfully, most high-profile journals now require submission of code and detailed descriptions of reported methods, thus paving a path toward increased transparency and shared access.

It is also noteworthy that AI cancer models of today have a strong emphasis on image and omics data. But one of the richest data sources of patient health and clinical history is embedded in the electronic health records (EHR) of a patient and still remains hugely underutilized. Reasons for this include records being unstructured with high levels of noise, sparseness, and inconsistencies, requiring dedicated curation and data cleaning. These challenges are being actively addressed by standards such as the Observational Medical Outcomes Partnership Common Data Model, which is focused on restructuring patient data into easy-to-use databases with standardized disease codes and harmonized vocabulary. This is further aided by user-friendly software that allows visualization of longitudinal patient data (e.g., PatientExploreR) and frameworks that facilitate mining of EHR to make clinically relevant predictions (126, 127).

From the clinical perspective, building clinicians' trust in AI-assisted decision-making is also critical for the entry of AI in clinic. To this end, Begoli and colleagues recommend development and adoption of systematic and pragmatic measures of uncertainty quantification in AI models (128). Uncertainty in a model may come from the choice of data, accuracy and completeness of data, inherent biases in the data, artifacts, and model misspecifications. Estimation of uncertainty in data-driven prediction models is an area of active research and in the future will provide a systematic framework for improving models and increasing confidence in AI-assisted clinical decision-making. Deep learning currently has the

reputation of being a “black box” but is in essence capturing complex correlations within data. Hence, additional research to increase model interpretability by understanding how deep learning models learn from given data, and what cellular and molecular mechanistic insights such models can provide, will also make the clinical use of AI models more agreeable to clinicians.

Thinking prospectively, prevention rather than treatment may end up being the most compelling application of AI to cancer care. Seminal research has already led the community to compile a portfolio of risk factors for cancer. Advances in technology have enabled various means of collecting data at an individual patient level. Aside from genetic tests and EHR, sensors from smartphones or other wearable devices also collect vast amounts of data points just for a single patient. These data can empower AI to improve precision of diagnosis by sensing physiologic and environmental status. They may help facilitate highly personalized disease prevention and treatment plans for each patient. Such AI systems may help monitor patients with cancer remotely and alert clinicians if need be. In the future, AI models that integrate genetic predispositions and EHR, together with lifestyle and environmental factors, may be able to accurately assess cancer risk for a person nearly in real time and suggest personalized options for early intervention and appropriate management of risk factors.

Authors' Disclosures

C. Gilvary reports personal fees from OneThree Biotech outside the submitted work; in addition, C. Gilvary has a patent for 62/393481 pending. N.S. Madhukar is currently employed by OneThree Biotech and holds equity in the company. O. Elemento reports other from OneThree Biotech, grants, personal fees, and other from Volastra Therapeutics, other from Owkin, and other from Freenome during the conduct of the study. No disclosures were reported by the other author.

Acknowledgments

O. Elemento is supported by NIH grants UL1TR002384 and R01CA194547, and the Leukemia and Lymphoma Society Specialized Center of Research grants 180078-02 and 7021-20.

Received January 20, 2021; revised February 6, 2021; accepted February 8, 2021; published first April 2, 2021.

REFERENCES

- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199–210.
- Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajirasouliha I. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine* 2018;27:317–28.
- Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell* 2018;173:305–20.
- Cheerla A, Gevaert O. Deep learning with multimodal representation for pancreatic prognosis prediction. *Bioinformatics* 2019;35:i446–54.
- Madhukar NS, Khade PK, Huang L, Gayvert K, Galletti G, Stogniew M, et al. A Bayesian machine learning approach for drug target identification using diverse data types. *Nat Commun* 2019;10:5221.
- Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Arch Pathol Lab Med* 2019;143:859–68.
- Al-Haija QA, Adebajo A. Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In: *Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*; 2020 Sep 9–12; IEEE; 2020. p. 1–7.
- Li S, Jiang H, Pang W. Joint multiple fully connected convolutional neural network with extreme learning machine for hepatocellular carcinoma nuclei grading. *Comput Biol Med* 2017;84:156–67.
- Korbar B, Olofson AM, Mirafior AP, Nicka CM, Suriawinata MA, Torresani L, et al. Deep learning for classification of colorectal polyps on whole-slide images. *J Pathol Inform* 2017;8:30.
- Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559–67.
- Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci Rep* 2020;10:1504.
- Campanella G, Hanna MG, Geneslaw L, Mirafior A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;25:1301–9.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
- Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 2016;35:1207–16.
- Jiang Y, Liang X, Wang W, Chen C, Yuan Q, Zhang X, et al. Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning. *JAMA Netw Open* 2021;4:e2032269.
- Wang X, Yang W, Weinreb J, Han J, Li Q, Kong X, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. *Sci Rep* 2017;7:15415.
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafi H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89–94.
- Freeman K, Dinnes J, Chuchu N, Takwoingi Y, Bayliss SE, Martin RN, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ* 2020;368:m127.
- Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht UV. mHealth 2.0: experiences, possibilities, and perspectives. *JMIR Mhealth Uhealth* 2014;2:e24.
- Nagpal K, Foote D, Liu Y, Chen PC, Wulczyn E, Tan F, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med* 2019;2:48.
- Zhou Q, Zhou Z, Chen C, Fan G, Chen G, Heng H, et al. Grading of hepatocellular carcinoma using 3D SE-DenseNet in dynamic enhanced MR images. *Comput Biol Med* 2019;107:47–57.
- Grewal JK, Tessier-Cloutier B, Jones M, Gakkhar S, Ma Y, Moore R, et al. Application of a neural network whole transcriptome-based pan-cancer method for diagnosis of primary and metastatic cancers. *JAMA Netw Open* 2019;2:e192597.
- Mostavi M, Chiu YC, Huang Y, Chen Y. Convolutional neural network models for cancer type prediction based on gene expression. *BMC Med Genomics* 2020;13:44.
- Penson A, Camacho N, Zheng Y, Varghese AM, Al-Ahmadie H, Razavi P, et al. Development of genome-derived tumor type prediction to inform clinical cancer care. *JAMA Oncol* 2019;6:84–91.
- Ringnér M, Peterson C. Microarray-based cancer diagnosis with artificial neural networks. *BioTechniques* 2003;34:S30–5.

27. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. *Nature* 2018;555:469–74.
28. Sun Y, Zhu S, Ma K, Liu W, Yue Y, Hu G, et al. Identification of 12 cancer types through genome deep learning. *Sci Rep* 2019;9:17256.
29. Moss EL, Gorsia DN, Collins A, Sandhu P, Foreman N, Gore A, et al. Utility of circulating tumor DNA for detection and monitoring of endometrial cancer recurrence and progression. *Cancers* 2020;12:2231.
30. Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, et al. Integrating genomic features for non-invasive early lung cancer detection. *Nature* 2020;580:245–51.
31. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 2018;10:eaat4921.
32. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018;359:926–30.
33. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* 2018;36:983–87.
34. Park H, Chun SM, Shim J, Oh JH, Cho EJ, Hwang HS, et al. Detection of chromosome structural variation by targeted next-generation sequencing and a deep learning application. *Sci Rep* 2019;9:3644.
35. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* 2019;53:1800986.
36. Mu W, Jiang L, Zhang J, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* 2020;11:5228.
37. Shboul ZA, Chen J, Iftekharuddin KM. Prediction of molecular mutations in diffuse low-grade gliomas using MR imaging features. *Sci Rep* 2020;10:3711.
38. Chen M, Zhang B, Topatana W, Cao J, Zhu H, Juengpanich S, et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *NPJ Precis Oncol* 2020;4:14.
39. Sidaway P. MSI-H: a truly agnostic biomarker? *Nat Rev Clin Oncol* 2020;17:68.
40. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019;25:1054–56.
41. Yamashita R, Long J, Longacre T, Peng L, Berry G, Martin B, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol* 2021;22:132–41.
42. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* 2019;30:44–56.
43. Krøigård AB, Thomassen M, Lænkholm AV, Kruse TA, Larsen MJ. Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. *PLoS One* 2016;11:e0151664.
44. Jain MS, Massoud TF. Predicting tumour mutational burden from histopathological images using multiscale deep learning. *Nat Mach Intell* 2020;2:356–62.
45. Wang L, Jiao Y, Qiao Y, Zeng N, Yu R. A novel approach combined transfer learning and deep learning to predict TMB from histology image. *Pattern Recognit. Lett* 2020;135:244–8.
46. He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer* 2020;8:e000550.
47. Xu Z, Verma A, Naveed U, Bakhoum S, Khosravi P, Elemento O. Using histopathology images to predict chromosomal instability in breast cancer: a deep learning approach. *medRxiv*.
48. Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R, et al. Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon Research Institute. *J Clin Oncol* 2013;31:217–23.
49. Greco FA. Molecular diagnosis of the tissue of origin in cancer of unknown primary site: useful in patient management. *Curr Treat Options Oncol* 2013;14:634–42.
50. Jiao W, Atwal G, Polak P, Karlic R, Cuppen E; PCAWG Tumor Subtypes and Clinical Translation Working Group. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. *Nat Commun* 2020;11:728.
51. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat Genet* 2018;50:1171–9.
52. Hoffman GE, Bendt J, Girdhar K, Schadt EE, Roussos P. Functional interpretation of genetic variants using deep learning predicts impact on chromatin accessibility and histone modification. *Nucleic Acids Res* 2019;47:10597–611.
53. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med* 2011;3:75ra26.
54. Bai H, Wang Z, Chen K, Zhao J, Lee JJ, Wang S, et al. Influence of chemotherapy on EGFR mutation status among patients with non-small-cell lung cancer. *J Clin Oncol* 2012;30:3077–83.
55. Haider S, Tyekucheva S, Prandi D, Fox NS, Ahn J, Xu AW, et al. Systematic assessment of tumor purity and its clinical implications. *JCO Precis Oncol* 2020;4:PO.20.00016.
56. Akbar S, Peikari M, Salama S, Panah AY, Nofech-Mozes S, Martel AL. Automated and manual quantification of tumour cellularity in digital slides for tumour burden assessment. *Sci Rep* 2019;9:14099.
57. Bhinder B, Elemento O. Computational methods in tumor immunology. *Methods Enzymol* 2020;636:209–59.
58. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep* 2018;23:181–93.
59. Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol* 2018;31:1502–12.
60. Fassler DJ, Abousamra S, Gupta R, Chen C, Zhao M, Paredes D, et al. Deep learning-based image analysis methods for brightfield-acquired multiplex immunohistochemistry images. *Diagn Pathol* 2020;15:100.
61. Torroja C, Sanchez-Cabo F. Digitaldsorter: deep-learning on scRNA-Seq to deconvolute gene expression data. *Front Genet* 2019;10:978.
62. Menden K, Marouf M, Oller S, Dalmia A, Magruder DS, Kloiber K, et al. Deep learning-based cell composition analysis from tissue expression profiles. *Sci Adv* 2020;6:eaba2619.
63. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas. *Elife* 2017;6:e27041.
64. Finotello F, Rieder D, Hackl H, Trajanoski Z. Next-generation computational tools for interrogating cancer immunity. *Nat Rev Genet* 2019;20:724–46.
65. Tran NH, Qiao R, Xin L, Chen X, Shan B, Li M. Personalized deep learning of individual immunopeptidomes to identify neoantigens for cancer vaccines. *Nat Mach Intell* 2020;2:764–71.
66. Tong Z, Zhou Y, Wang J. Identifying potential drug targets in hepatocellular carcinoma based on network analysis and one-class support vector machine. *Sci Rep* 2019;9:10442.
67. López-Cortés A, Paz-Y-Miño C, Cabrera-Andrade A, Barigye SJ, Munteanu CR, González-Díaz H, et al. Gene prioritization, communality analysis, networking and metabolic integrated pathway to better understand breast cancer pathogenesis. *Sci Rep* 2018;8:16679.
68. López-Cortés A, Cabrera-Andrade A, Vázquez-Naya JM, Pazos A, González-Díaz H, Paz-Y-Miño C, et al. Prediction of breast cancer proteins involved in immunotherapy, metastasis, and RNA-binding

- using molecular descriptors and artificial neural networks. *Sci Rep* 2020;10:8515.
69. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, et al. Integrating genotype and phenotype information: an overview of the PharmGKB project. *Pharmacogenetics Research Network and Knowledge Base. Pharmacogenomics J* 2001;1:167–70.
 70. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* 2018;10:25.
 71. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell* 2017;170:564–76.
 72. Gilvary C, Madhukar NS, Gayvert K, Foronda M, Perez A, Leslie CS, et al. A machine learning approach predicts essential genes and pharmacological targets in cancer. *bioRxiv*.
 73. Chen MM, Li J, Mills GB, Liang H. Predicting cancer cell line dependencies from the protein expression data of reverse-phase protein arrays. *JCO Clin Cancer Inform* 2020;4:357–66.
 74. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. *J Cheminform* 2017;9:48.
 75. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv* 2018;4:eaap7885.
 76. You J, Liu B, Ying R, Pande V, Leskovec J. Graph convolutional policy network for goal-directed molecular graph generation. *arXiv*.
 77. Maziarka L, Pocha A, Kaczmarczyk J, Rataj K, Danel T, Warchol M. Mol-CycleGAN: a generative model for molecular optimization. *J Cheminform* 2020;12:2.
 78. De Cao N, Kipf T. MolGAN: an implicit generative model for small molecular graphs. *arXiv*.
 79. Gayvert KM, Madhukar NS, Elemento O. A data-driven approach to predicting successes and failures of clinical trials. *Cell Chem Biol* 2016;23:1294–301.
 80. Martins IF, Teixeira AL, Pinheiro L, Falcao AO. A Bayesian approach to in silico blood-brain barrier penetration modeling. *J Chem Inf Model* 2012;52:1686–97.
 81. Shen J, Cheng F, Xu Y, Li W, Tang Y. Estimation of ADME properties with substructure pattern recognition. *J Chem Inf Model* 2010;50:1034–41.
 82. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;171:1437–52.
 83. Mokou M, Lygirou V, Angelioudaki I, Paschalidis N, Stroggilos R, Frantzi M, et al. A novel pipeline for drug repurposing for bladder cancer based on patients' omics signatures. *Cancers* 2020;12:3519.
 84. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun* 2017;8:16022.
 85. Mastrogamvraki N, Zaravinos A. Signatures of co-deregulated genes and their transcriptional regulators in colorectal cancer. *NPJ Syst Biol Appl* 2020;6:23.
 86. Li B, Dai C, Wang L, Deng H, Li Y, Guan Z, et al. A novel drug repurposing approach for non-small cell lung cancer using deep learning. *PLoS One* 2020;15:e0233112.
 87. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;41:D955–61.
 88. Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer* 2020;1:235–48.
 89. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;6:813–23.
 90. Chang Y, Park H, Yang HJ, Lee S, Lee KY, Kim TS, et al. Cancer Drug Response Profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep* 2018;8:8857.
 91. Gilvary C, Elkhader J, Madhukar N, Henschcliffe C, Goncalves MD, Elemento O. A machine learning and network framework to discover new indications for small molecules. *PLoS Comput Biol* 2020;16:e1008098.
 92. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 2019;35:5191–8.
 93. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol* 2011;7:496.
 94. Liu D, Schilling B, Liu D, Sucker A, Livingstone E, Jerby-Arnon L, et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat Med* 2019;25:1916–27.
 95. Litchfield K, Reading JL, Puttick C, Thakkar K, Abbosh C, Bentham R, et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 2021;184:596–614.
 96. Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y, et al. Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. *Clin Cancer Res* 2021;27:131–40.
 97. Sun D, Wang M, Li A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans Comput Biol Bioinform* 2019;16:841–50.
 98. Hao J, Kim Y, Kim TK, Kang M. PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *BMC Bioinformatics* 2018;19:510.
 99. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res* 2018;24:1248–59.
 100. Korfiatis P, Kline TL, Lachance DH, Parney IF, Buckner JC, Erickson BJ. Residual deep convolutional neural network predicts MGMT methylation status. *J Digit Imaging* 2017;30:622–8.
 101. Kim H, Goo JM, Lee KH, Kim YT, Park CM. Preoperative CT-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas. *Radiology* 2020;296:216–24.
 102. Lou B, Doken S, Zhuang T, Wingerter D, Gidwani M, Mistry N, et al. An image-based deep learning framework for individualizing radiotherapy dose. *Lancet Digit Health* 2019;1:e136–47.
 103. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 2018;115:E2970–9.
 104. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018;8:3395.
 105. Skrede OJ, De Raedt S, Kleppe A, Hveem TS, Liestøl K, Maddison J, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020;395:350–60.
 106. Ferrara R, Matos I. Atypical patterns of response and progression in the era of immunotherapy combinations. *Future Oncol* 2020;16:1707–13.
 107. Dercle L, Fronheiser M, Lu L, Du S, Hayes W, Leung DK, et al. Identification of non-small cell lung cancer sensitive to systemic cancer therapies using radiomics. *Clin Cancer Res* 2020;26:2151–62.
 108. Choi JH, Kim HA, Kim W, Lim I, Lee I, Byun BH, et al. Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning. *Sci Rep* 2020;10:21149.
 109. Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 2019;25:3266–75.
 110. Blasiak A, Khong J, Kee T. CURATE.AI: optimizing personalized medicine with artificial intelligence. *SLAS Technol* 2020;25:95–105.
 111. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483:603–7.
 112. Smirnov P, Kofia V, Maru A, Freeman M, Ho C, El-Hachem N, et al. PharmacDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Res* 2018;46:D994–D1002.

113. Gillet JP, Varma S, Gottesman MM. The clinical relevance of cancer cell lines. *J Natl Cancer Inst* 2013;105:452–8.
114. Cheung ST, Chan SL, Lo KW. Contaminated and misidentified cell lines commonly use in cancer research. *Mol Carcinog* 2020;59:573–4.
115. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. *Cell* 2016;166:740–54.
116. Cortés-Ciriano I, van Westen GJ, Bouvier G, Nilges M, Overington JP, Bender A, et al. Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics* 2016;32:85–95.
117. Chiu YC, Chen HH, Zhang T, Zhang S, Gorthi A, Wang LJ, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics* 2019;12:18.
118. Sakellaropoulos T, Vougas K, Narang S, Koinis F, Kotsinas A, Polyzos A, et al. A deep learning framework for predicting response to therapy in cancer. *Cell Rep* 2019;29:3367–73.
119. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* 2020;38:672–84.
120. Menden MP, Wang D, Mason MJ, Szalai B, Bulusu KC, Guan Y, et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun* 2019;10:2674.
121. Gilvary C, Dry JR, Elemento O. Multi-task learning predicts drug combination synergy in cells and in the clinic. *bioRxiv*.
122. Yuan J, Hu Z, Mahal BA, Zhao SD, Kensler KH, Pi J, et al. Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer Cell* 2018;34:549–60.
123. Bhargava HK, Leo P, Elliott R, Janowczyk A, Whitney J, Gupta S, et al. Computationally derived image signature of stromal morphology is prognostic of prostate cancer recurrence following prostatectomy in African American patients. *Clin Cancer Res* 2020;26:1915–23.
124. Driehuis E, Kretzschmar K, Clevers H. Establishment of patient-derived cancer organoids for drug-screening applications. *Nat Protoc* 2020;15:3380–409.
125. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Massive Analysis Quality Control (MAQC) Society Board of Directors, Waldron L, et al. Transparency and reproducibility in artificial intelligence. *Nature* 2020;586:E14–6.
126. Glicksberg BS, Oskotsky B, Thangaraj PM, Giangreco N, Badgeley MA, Johnson KW, et al. PatientExploreR: an extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model. *Bioinformatics* 2019;35:4515–8.
127. Landi I, Glicksberg BS, Lee HC, Cherng S, Landi G, Danieleto M, et al. Deep representation learning of electronic health records to unlock patient stratification at scale. *NPJ Digit Med* 2020;3:96.
128. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat Mach Intell* 2019;1:20–3.