

# Discovery of predictive biomarkers for cancer therapy through computational approaches

Xin Wang <sup>1,13</sup>, Julia Nguyen <sup>1,2,13</sup>, Kristen Nader <sup>3,4</sup>, Mitro Miihkinen <sup>3,4</sup>, Patrick Wall<sup>5</sup>, Akshat Singhal <sup>6,7</sup>, Philippe L. Bedard<sup>1</sup>, Trey Ideker<sup>5,6,7,8</sup> , Tero Aittokallio <sup>3,4,9,10</sup>  & Benjamin Haibe-Kains<sup>2,11,12</sup> 

## Abstract

Precision oncology involves the use of predictive biomarkers to personalize treatment. However, for most cancer therapeutics or combination regimens, effective biomarkers have been elusive. This challenge has fuelled efforts to interrogate increasingly diverse and complex clinical and molecular determinants of treatment response. Some molecular predictors have been identified (for example, based on analysis of transcriptomic or imaging data), although the limited reproducibility and robustness of many of these candidate biomarkers make them difficult to apply in clinical practice. Moreover, different types of predictor must often be combined to optimize treatment selection (for example, gene signatures plus patient characteristics). Computational methods, including machine learning and artificial intelligence approaches, provide opportunities to identify predictive patterns in both clinical data and preclinical datasets and to predict treatment response for individual patients. Such approaches also offer opportunities to predict the efficacy or synergy of drug combinations, for example, via extrapolation from correlations of monotherapy responses or by linking the cellular responses observed in preclinical drug screens with molecular and clinical data from patients. In this Review, we describe the application of computational methods to predictive biomarker discovery, including current progress, key challenges facing this field, and future opportunities.

## Sections

Introduction

Sources of data for biomarker discovery


Monotherapy biomarker discovery

Combination therapy biomarkers

Open computational challenges

Translational challenges

Conclusions

A full list of affiliations appears at the end of the paper.  e-mail: [tideker@health.ucsd.edu](mailto:tideker@health.ucsd.edu); [tero.aittokallio@helsinki.fi](mailto:tero.aittokallio@helsinki.fi); [benjamin.haibe-kains@uhn.ca](mailto:benjamin.haibe-kains@uhn.ca)

## Key points

- The discovery and validation of predictive biomarkers is an essential prerequisite to advances in personalized cancer therapies.
- Predictive biomarkers can be derived from analyses of molecular, radiological and/or histopathological data as well as clinical data, and multiple modalities can be combined to capture more complex biological associations.
- Both preclinical models and patient data are useful sources for the discovery of predictive biomarkers, and several publicly available collections exist that can also support biomarker investigations.
- Computational methods and machine learning algorithms can be used to predict responses to treatment both with monotherapies and combination regimens.
- Several remaining challenges relating to the discovery of predictive biomarkers must be addressed in order to improve the generalizability of predictive models, to better capture tumour heterogeneity and to overcome the various limitations that can hinder clinical translation.

## Introduction

Precision oncology, in which therapies are tailored to the tumour type, microenvironment, cellular and/or molecular fingerprints of cancer in each patient, has shifted from a bold aspiration to an urgent clinical necessity<sup>1</sup>. The advent of targeted therapies and immunotherapies has revolutionized oncology and offers the promise of precision over empiricism. Predictive biomarkers, which provide molecular clues that can help in foreseeing how a patient will respond to a given treatment, lie at the heart of this revolution<sup>2</sup>. From circulating DNA fragments in the blood to patterns that are not apparent to the human eye in radiological images or histopathological slides, these biomarkers now span an expanding spectrum of biological, clinical and digital domains. Despite their promise, the search for robust and reproducible biomarkers remains elusive, hampered by the inherent biological complexity of cancer and the resultant resistance to reductionist approaches. This paradox is driving a new wave of discovery powered by machine learning and artificial intelligence that aims not only to provide a more detailed understanding of tumour biology but also to translate this information into actionable predictions that can reshape the way we treat patients.

Molecularly targeted therapies were developed in a paradigm that assumes a direct link between a genetic alteration, its protein product and sensitivity to a matched targeted therapeutic agent. Successful examples of targeted therapies include tropomyosin receptor kinase inhibitors for tumours harbouring *NTRK* fusions<sup>3</sup> and HER2-directed therapies for those with *HER2* overexpression<sup>4</sup>, among others. Between 2000 and 2022, the FDA approved 573 agents for various oncological indications, yet among these, only 38.9% were approved for a biomarker-defined population<sup>5</sup>. This paucity of approved biomarkers has led to widespread adoption of next-generation sequencing panels and the current model of genotype-matched clinical trials, in which the presence of specific genomic aberrations can be predictive of response to specific agents<sup>6–8</sup>. While most current clinically approved biomarkers are univariable (focused on a single molecular

feature), this traditional framework has limitations in addressing the complexity of cancer biology. Furthermore, few biomarkers to date have incorporated features derived from the complex spatial and temporal heterogeneity of many cancers<sup>9,10</sup>. Importantly, biomarkers based on the presence of drug targets alone are often poor indicators of a clinical response<sup>11–13</sup> and are often not predictive of sensitivity to combination regimens. Consequently, computational approaches are poised to transform the current biomarker landscape by enabling the discovery of multivariable and multimodal biomarkers capable of predicting treatment responses across diverse therapeutic contexts.

The explosion of high-throughput omics technologies, encompassing (epi)genomics, transcriptomics, proteomics and metabolomics, has generated a deluge of multidimensional data from diverse cancer types and clinical contexts. These datasets, coupled with those from preclinical models, such as cell lines, patient-derived organoids (PDOs) and patient-derived xenografts (PDXs), provide ample opportunity for computational exploration and experimental investigation of predictive biomarkers. Artificial intelligence and machine learning algorithms, including supervised, unsupervised and deep learning models<sup>14,15</sup>, have demonstrated the capacity to identify subtle patterns within these large-scale and heterogeneous datasets that elude traditional analytical methods. An important advantage of artificial intelligence and machine learning models is the ability to identify predictive biomarker candidates without the need for a detailed understanding of the underlying mechanisms of action<sup>16</sup>; this general principle remains relevant regardless of whether the models are applied to the identification of predictive biomarkers for monotherapies or combination regimens<sup>17</sup>, although biological validation of candidate biomarkers remains a crucial step. Successful deployment of these methods will require ever larger and more specific clinical datasets for training and prospective clinical validation to mitigate the risks of overfitting and a lack of generalizability. Predictive biomarkers are not static entities, but rather dynamic tools that evolve along a patient's treatment journey, from diagnosis (for first-line treatment selection) through possible remission (for minimal residual disease assessments) to guiding decisions at relapse or disease progression (for later line treatment selection)<sup>18,19</sup>. As such, the clinical relevance of biomarkers can change over time, with distinct biomarkers often becoming informative at different disease stages. This dynamic nature highlights the critical role of biomarkers in tailoring therapy across the continuum of disease. The current biomarker landscape provides great opportunities for computational biomarker discovery, although challenges also remain, both in terms of technological limitations as well as those relating to the clinical implementation of longitudinal biomarkers in light of the exponential resources required for serial testing in routine clinical settings<sup>20</sup>.

In this Review, we provide a narrative discussion of the computational discovery of predictive biomarkers in oncology, referencing systematic reviews and benchmarking studies to contextualize recent advances. We first examine the most widely used clinical and omics data available for biomarker discovery along with common preclinical model systems. Subsequently, we explore primary computational approaches to biomarker discovery both in the context of monotherapy or combination regimens, followed by a discussion of the major unsolved computational challenges. This Review aims to provide an integrative perspective on the opportunities and challenges that lie ahead, paving the way for innovative solutions that address the complexities of tumour biology and drug responses.

## Sources of data for biomarker discovery

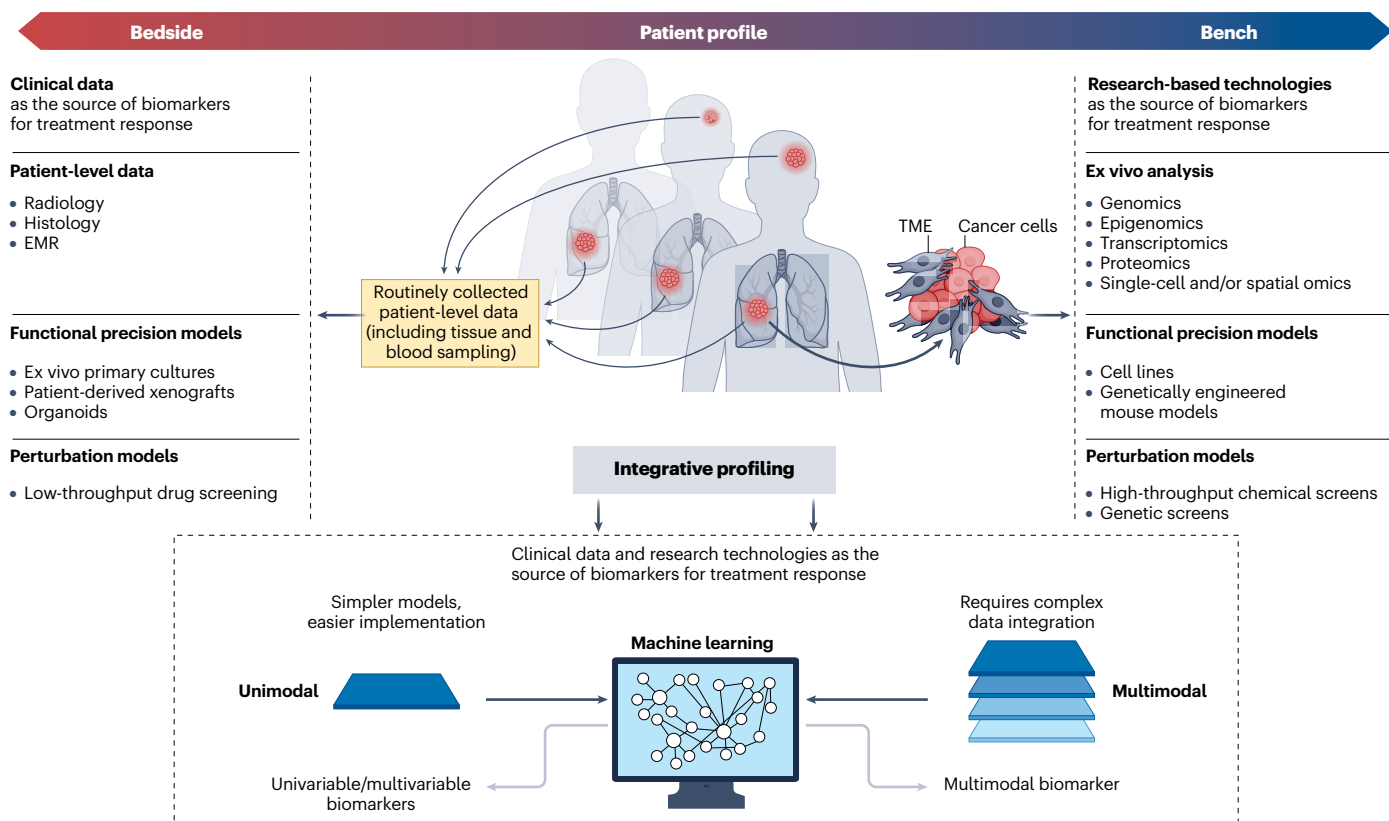
### Clinical and omics data

Molecular, radiological, histopathological and observational data are routinely screened for biological factors that might inform the detection of cancer, provide a more accurate prognosis and/or guide clinical management (Fig. 1). Improvements in the resolution and sensitivity of these profiling technologies as well as the emergence of new data modalities provide highly detailed descriptions of tumours that can be leveraged to identify biomarkers across biological scales. Within the field of precision oncology, differentiating predictive biomarkers, which are indicators of treatment response, from prognostic biomarkers, which provide insights into the risks of disease progression irrespective of therapy, is an important starting point<sup>2</sup>.

The foundations for cancer treatment are provided by observational data accessed through clinical records. Data regarding patient demographics (for example, age, sex and ethnicity), lifestyle (such as smoking history and diet), family history of cancer, the results of relevant laboratory tests and comorbidities are aggregated into a comprehensive summary of relevant patient profiles<sup>21</sup>. These profiles also include histopathological information regarding tumour grade, subtype and invasion patterns identified in tumour tissues.

Such clinical information contextualizes specific cancers with respect to similar cases, which have become easier to cross-reference through the digitalization of profiles into electronic medical records (EMRs). Additionally, clinical data provide a framework through which to analyse more detailed information on molecular characteristics and/or imaging. Efforts to date have required expert curation, although natural language processing is poised to transform this field and broaden the use of routinely collected clinical data for biomarker prediction<sup>22</sup>.

Imaging data collected from radiography provides a macroscopic view of tumours in their environment. Similarly, immunohistological staining of histopathology slides can provide morphological information in addition to the location and abundance of protein markers such as HER2, overexpression of which in breast cancer samples is predictive of tumour response to anti-HER2 agents<sup>23</sup>. Molecular data, encompassing a wide range of modalities, including genomics, epigenomics, proteomics, transcriptomics and metabolomics, are used to identify aberrations affecting the basic functions of the cellular machinery that drive the development and/or progression of cancer as well as affecting prognosis. Some of these alterations are also potentially targetable, for example, *BRAF*<sup>V600E</sup> mutations in patients with melanoma and, more recently, *KRAS*<sup>G12C</sup> in patients with colorectal cancer or non-small-cell



**Fig. 1 | Sources of predictive biomarkers based on patient profiling data.**

Left: clinical data routinely collected from patients, including tissue and/or liquid biopsies, electronic medical records (EMRs) and radiological and/or histopathological images. Functional precision models can be derived from patient tissue samples or low-passage dissociated cells to ensure they remain representative of patients. These models are typically subjected to low-throughput drug screening. Right: research-oriented data collection such as (epi)genomics, transcriptomics and proteomics sequencing using bulk, single-cell or spatial

profiling technologies. Functional precision models, including patient-derived or immortalized models, can be subjected to chemical screening (such as high-throughput drug screens) and genetic perturbations (such as CRISPR). Bottom: integrative profiling of both clinical and research-oriented data can provide new sources of biomarkers. Computational methods, such as machine learning, can identify univariable or multivariable biomarkers from single input modalities or multimodal biomarkers from the integration of multiple input modalities (multimodal). TME, tumour microenvironment.

**Table 1 | Advantages and disadvantages of the different data modalities for biomarker discovery**

Modality	Advantages	Disadvantages	Extent of clinical implementation
<b>Clinical data</b>			
Electronic medical records	Data can be obtained longitudinally and stored in centralized registries	Unstructured, lacking in standardized recording systems across different registries	Universal
Histology	Provides clinically relevant information on the expression of specific biomarkers, including spatial location	Provides limited molecular insight beyond specific biomarkers and/or histologies, interpretation can be subjective. Analysis of biopsy material might provide biased results	Universal
Imaging	Provides non-invasive direct visualization of tumours, amenable to longitudinal assessments	Provides limited additional molecular insights	Universal
<b>Omics analysis</b>			
DNA sequencing	Robust, reliable and cost-effective relative to most other molecular analyses	Use of targeted panels or assays provides limited information beyond the presence or absence of specific alterations, which might be limited in certain samples. Can also fail to fully capture tumour heterogeneity	Widely used
Epigenomics	Epigenetic modifications are more widely observed across the genome (compared with mutations)	Data are often difficult to interpret, which is made more challenging by the need for fast turnaround in clinical settings	Implemented in certain specific settings
Transcriptomics	Provides a rich source of context-specific information	Assays can be noisy, unstable and time-dependent	Implemented in certain specific settings
Proteomics	Enables detailed low-throughput investigations of specific proteins (via immunohistochemistry and related techniques)	Not amenable to high-throughput investigations of large numbers of proteins	Widely implemented for a single or few proteins
Single-cell omics	Enables assessments of tumour heterogeneity as well as limited-abundance cell types	Expensive and complex to implement, particularly within clinically relevant turnaround times	Not clinically implemented
Spatial omics	Adds a further level of spatial information to omics analysis	Expensive, computationally intensive and complex to implement, particularly within clinically relevant turnaround times	Not clinically implemented

lung cancer (NSCLC)<sup>24</sup>. Technological advances have enabled single-cell and spatial multi-omics profiling<sup>25</sup>, which has the potential to highlight intracellular molecular features that can be associated with treatment response<sup>26,27</sup>. Liquid biopsies have also garnered interest as a method of identifying molecular biomarkers present in circulating tumour DNA without the need for solid tumour biopsy<sup>28</sup>.

Whereas extensive investigations of data obtained using each of these modalities have revealed predictive biomarkers, the combination of multiple data modalities can provide a more comprehensive representation of the biological interactions governing cancer heterogeneity and overcome the limitations of individual data modalities (Table 1). Various studies have combined data from similar units and/or scales, such as integrating radiological and pathological imaging data to predict chemotherapy response in patients with rectal cancer<sup>29</sup>, as well as combining data from different scales, for example, combining information on clinical characteristics and genomic alterations with CT scans and digitized histopathology slides to predict responsiveness to immune-checkpoint inhibitors (ICIs) in patients with NSCLC<sup>30</sup>. For the field to continue leveraging these multimodal approaches, limitations in data availability must be addressed to ensure robust biomarker discovery, reproducibility and feasibility of clinical deployment.

## Preclinical model systems

Preclinical model systems, including cancer cell lines or primary cell cultures, PDOs and PDXs, are invaluable tools for understanding disease mechanisms in controlled environments and provide a strong rationale for leveraging the abundance of preclinical data for computational

biomarker discovery<sup>31–33</sup>. Another major advantage of model systems is the ability to overcome the challenges related to clinical data availability (Box 1) to enable biomarker discovery using computational methods that demand large sample sizes. Although the richness of preclinical data has the potential to enable the development of sophisticated computational models, clinical companion diagnostics have yet to routinely leverage preclinical data for drug response predictions. Preclinical models are unable to capture the complexity of human tumours, and harmonizing quality data across preclinical and clinical data domains remains an ongoing challenge. The inherent ability of artificial intelligence and machine learning approaches to prioritize informative patterns from complex data can mitigate these weaknesses and support the integration of data from different preclinical sources. For example, multi-omics assessments can identify models that are most representative of their specific tumour subtypes to improve the translatability of biomarker discovery<sup>34–37</sup>. Tools such as Celligner<sup>38</sup> and PRECISE<sup>39</sup> can also be used to identify preclinical models with molecular features that most closely align with those of clinical samples and/or improve the alignment between preclinical and clinical feature spaces. Building highly predictive models using a combination of preclinical and clinical data is also possible with certain machine learning architectures, for example, by using few-shot learning to translate predictions from high-throughput screens to individual patients<sup>40</sup>. Despite these opportunities, preclinical data remain heavily underused in biomarker development, warranting discussions on how to develop and define better preclinical models, address the computational challenges to maximize the translational potential of preclinical model systems and

leverage this abundant source of data for the computational discovery of clinically relevant biomarkers. To expedite preclinical biomarker discovery, a broad array of open-access and large-scale preclinical data resources has already been established<sup>41–45</sup>. Although these resources offer standardized datasets encompassing various layers of information, including molecular profiles and drug interactions, the development of more representative preclinical models will be necessary to produce more faithful predictions of the outcomes of patients.

In addition to the use of preclinical model systems for computational biomarker discovery, advances in functional profiling technologies, such as single-cell sequencing, are poised to further improve both the scale and precision of preclinical resources in the near future<sup>46–48</sup>. *Ex vivo* drug testing in culture models and PDOs can provide high-resolution response data<sup>49</sup>, supporting the discovery of predictive functional biomarkers along with other pharmaceutical applications, such as drug repurposing<sup>50,51</sup>, and providing biomarker-guided treatment options when standard-of-care approaches have been exhausted<sup>52</sup>. Both strategies can be leveraged to create a more tailored and faster approach to treatment selection and clinical testing<sup>53</sup>. However, *ex vivo* response profiling has shown the greatest clinical promise so far in haematological malignancies<sup>54,55</sup>, partly owing to the relative ease of developing cellular assays. Similar approaches, such as patient-specific ‘avatars’, are increasingly being developed and used in solid tumours<sup>56</sup>. Nevertheless, considerable work remains to be conducted to scale these methods for clinical practice.

## Monotherapy biomarker discovery

Despite the discovery and development of clinically actionable biomarkers that predict responses to monotherapies using traditional methods, the advent of artificial intelligence and machine learning has enabled development of the more sophisticated tools necessary for tackling complex mechanisms of action and/or phenomena such as tumour heterogeneity (Table 2 and Fig. 2). Nonetheless, at the time of writing, only a few computationally discovered predictive biomarkers have been successfully adopted or are in the process of translation into clinical use (Table 3). Hence, we discuss the challenges and barriers that hinder clinical implementation.

## Identifying features and response variables

Predictive computational models designed to identify biomarkers associated with response typically require a profile of explicit ‘machine-readable’ features as input. Observational clinical data provide a rich source of features relevant to predicting treatment response, and advances in artificial intelligence and machine learning have enabled automated extraction of the most relevant clinical features from EMRs<sup>22,57</sup>. Features can also be extracted from medical images (such as tumour volumes quantified using MRI or morphological features determined from pixel values of digitized haematoxylin-and-eosin-stained slides). Individual features obtained from molecular profiles can include genetic aberrations (such as mutations, fusions or copy-number variations), molecular characteristics (such as gene or protein expression, or metabolite levels), or various other characteristics (such as methylation status or chromatin accessibility at specific loci).

Multiple features can be combined to create a quantifiable meta-feature, often referred to as a signature. Signatures comprise a fixed number of features that can be mathematically combined using linear methods (for example, by applying weighted regression techniques or unsupervised clustering methods such as principal

component analysis or non-negative matrix factorization) or artificial intelligence or machine learning approaches, or simply selected as a panel, to calculate a signature score – a single value quantifying the extent of the signature within a patient. Tumour mutational burden (TMB), which reflects the number of mutations within the genome of cancer cells, is an example of a composite signature that can be used clinically as a predictor of responsiveness to ICIs<sup>58–60</sup>. Other signatures used in clinical practice to guide treatment selection in specific settings include microsatellite instability to guide the use of ICIs and homologous recombination deficiency for certain poly(ADP-ribose) polymerase (PARP) inhibitors in specific settings<sup>61–63</sup>. Although not in clinical use, several other molecular signatures have demonstrated associations with specific cancer aetiologies<sup>64–67</sup> such as the well-established COSMIC mutational signatures<sup>68,69</sup>. These signatures provide a promising direction for advancing precision oncology by capturing complex biological patterns while remaining interpretable.

In addition to specific input features, predictive models also require explicit prediction tasks. For biomarker discovery, this task is often the prediction of one or more treatment outcome variables. These outcomes can be binary (such as responder and non-responder), categorical (RECIST-defined response), proportional values (overall response rate (ORR) or clinical benefit rate), or time-to-event outcomes (overall survival or progression-free survival (PFS)). The choice of treatment response metric for model prediction must therefore consider the treatment type and the most appropriate outcomes for the specific clinical setting. For most immunotherapies, which can have non-linear response patterns and might provide long-term clinical benefit despite minimal evidence of tumour shrinkage in certain

## Box 1 | Clinical data – what, why and how

With the end goal of clinical deployment, the performance of newly discovered biomarkers must be rigorously assessed using large, high-quality clinical datasets. Clinical trials provide a rich source of such data because imaging, molecular profiling and response monitoring are often all mandated as part of the trial protocol. Nonetheless, although these data are often already stored, many research centres do not make patient-level data easily accessible, and even when data are shared, incomplete reporting can substantially limit their utility for predictive biomarker research. The selectivity and controlled settings of most clinical trials can also reduce the extent to which the data reflect the outcomes of real-world patients; hence, real-world data (RWD) have become an invaluable source of data for biomarker validation<sup>240</sup>. However, RWD availability is generally more limited than data from clinical trials, and the lack of standardization challenges the creation of RWD datasets that enable robust evaluations of effectiveness.

Datasets from patients receiving immunotherapies, mostly from clinical trials, have been widely released in recent years and have fuelled the discovery and large-scale validation of clinically actionable immune-related biomarkers<sup>241</sup>. Increased efforts to collect and share clinical data on the activity of additional cancer therapies across diverse patient populations will be required to continue advancing the field. Furthermore, establishing a standardized ‘minimum level of information’ would ensure the comprehensiveness of all shared datasets and thus maximize their utility for biomarker research.

**Table 2 | Comparison of computational methods for clinically oriented biomarker discovery**

Method	Interpretability	Sample size and data requirements	Computational and implementation complexity	Clinical challenges and extent of translation
Univariable statistical associations	Interpretable by nature; involves direct relationships; accessible to non-technical audiences	Requires simple, structured data; sample sizes range from dozens to hundreds	Requires only standard computational resources, minimal calculation times and minimal expertise	Limited predictive power, might oversimplify complex conditions and does not capture non-linear relationships; widely used in clinical research
Regression analyses	Interpretable by nature; coefficients easy to understand, with confidence intervals providing an indication of predictive uncertainty	Requires structured data with defined outcomes; sample sizes of hundreds typically required	More complex than statistical associations, although still feasible with standard computational resources equipped with standard statistical software packages; some expertise needed	Challenges include translating statistical findings into operational changes; might also require customization for specific populations; widely used in clinical research and for biomarker-based precision oncology
Classical artificial intelligence/machine learning algorithms	Requires additional methods for explainability; tree-based methods more interpretable; can use feature importance analysis	Can handle larger feature spaces and multiple modalities; some models can handle missing values well; requires sample sizes of hundreds to thousands	Requires moderate computational power and minutes to hours for training; data science experience needed	Requires validation across diverse populations; other challenges relate to clinician scepticism and interpretability; not translated to the clinic
Deep learning	Requires additional methods for explainability, owing to inherent 'black box' problem	Can handle larger feature spaces and multiple modalities, including complex unstructured data, images, text and/or time series; requires sample sizes of thousands to millions, potentially with the need for data augmentation	Requires substantial computational resources, including GPU acceleration and days to weeks for training; requires high levels of experience, including ongoing maintenance	Clinician scepticism owing to challenges with interpretability, 'black box'-related regulatory concerns and requires large datasets; not translated to the clinic
Foundation models <sup>a</sup>	Requires additional methods for explainability owing to inherent 'black box' problem	Can handle large feature spaces and multiple modalities; requires millions to billions of samples for training and hundreds to thousands for fine-tuning	Requires substantial computational resources, including GPU/TPUs as well as extensive parameter optimization, potentially requiring weeks to months of training; requires a specialist artificial intelligence/machine learning team	High costs, hallucination risks, model decay and the need for ongoing retraining to incorporate new data in real time all challenge development and regulatory approval process; not translated to the clinic

GPU, graphics processing unit; TPU, tensor processing unit. <sup>a</sup>Has the notable advantage of flexibility in the range of potential predictive tasks.

scenarios<sup>70–72</sup>, metrics that account for stable disease and/or long-term benefit, such as overall survival, PFS and clinical benefit rate, might be the most suitable variables. Alternatively, ORR might be a more appropriate variable for models designed to predict benefit from targeted therapies. The selected response variable will also determine the type of model required to predict the specific data type (binary classification, continuous output prediction or time-to-event prediction).

### Computational methods for biomarker discovery

Associations between treatment response and individual features, including specific mutations or signatures, can be computed using various univariable statistical methods and regression models that consequently identify univariable biomarkers. Alternatively, multivariable approaches use mathematical modelling to simultaneously assess multiple individual features or signatures for their collective contribution to predictions of treatment response. For example, MammaPrint<sup>73</sup>, a clinically validated predictor of benefit from adjuvant chemotherapy in patients with breast cancer, uses a logistic regression classifier to model the expression of 70 marker genes previously identified using univariable association testing. This multivariable biomarker has demonstrated the ability to identify patients with high-risk, early-stage breast cancer who can nonetheless safely avoid chemotherapy based on their genomic risk score<sup>74</sup>. Another example is HRDetect, a regularized logistic regression model that uses six mutational signatures to predict sensitivity to PARP inhibitors in patients with breast cancer<sup>75</sup>.

Although HRDetect has demonstrated accuracy in retrospective assessments<sup>76,77</sup>, validation in a prospective trial is needed to further determine its clinical relevance.

Univariable biomarkers offer several advantages over more complex multivariable predictors. Their simplicity can accelerate the discovery and testing processes with reduced risks of bias and overfitting (that is, the scenario in which the predictive model can closely fit the training data but fails to generalize to new data). Clinical implementation also becomes more feasible as the biological relevance of single features is usually easier to explain to end users and more cost-effective to measure in patients. However, relying on predictive models based on single features might limit predictive power; hence, multivariable predictors are often required to capture the more complex biological interactions that govern treatment response or resistance to therapies with complex and/or incompletely understood mechanisms of action (such as ICIs). The choice of a mathematical model for multivariable biomarker discovery will have implications for sample size requirements, interpretability and clinical applicability, warranting careful consideration (Table 2 and Fig. 3).

**Univariable statistical associations.** For continuous features and response variables, univariable associations can be quantified using statistical measures of similarity such as correlation coefficients, concordance index, cosine similarity and various distance measurements<sup>78</sup>. Statistical tests, including Wilcoxon rank sum test, Fisher's exact test,

# Review article

hazard ratios and logrank tests, can also be used to identify biomarkers by comparing biomarker levels between groups (such as responders versus non-responders), or by comparing response and/or survival outcomes between groups of patients stratified by the candidate biomarker (high versus low biomarker expression)<sup>79</sup>. Parametric assumptions and the data types being assessed are both important considerations when selecting the most appropriate statistical method.

**Regression analyses.** Linear or logistic regression models enable the simultaneous analysis of multiple independent variables while also allowing the quantification of the predictive contribution of each individual feature, and regularization techniques, such as LASSO, Ridge and Elastic net, can help mitigate overfitting. Regression models are also commonly used for univariable biomarker discovery to account for potential confounders, such as age, sex, tumour site of origin and/or histological grade, to ensure robust biomarker performance across diverse clinical factors and patient demographics<sup>80–82</sup>. The Cox proportional hazards model is a linear regression survival model that measures the associations between individual features and survival duration, such as overall survival and PFS, and has been used in the development of several clinically implemented multivariable signatures such as the Oncotype DX assay<sup>83</sup> among others<sup>84–86</sup>.

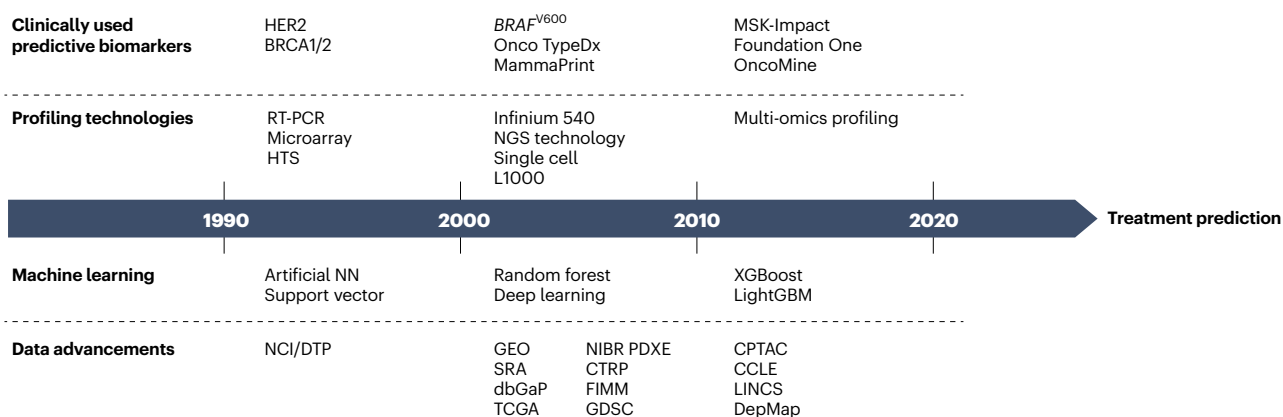
**Classical machine learning algorithms.** Alternative modelling techniques might be required to capture non-linear interactions between multiple independent variables and treatment response. Such predictive feature–response relationships can be captured using classical machine learning models such as random forests, gradient-boosting machines, Gaussian processes and support vector machines<sup>87–90</sup>. Several of these models have been adapted to more accurately model survival events such as random survival forests<sup>91</sup> and survival support vector machines<sup>92</sup>. An ongoing debate exists regarding whether linear or non-linear models are more effective as methods of drug response prediction<sup>13</sup>. Although non-linear models can capture more complex multivariate associations, understanding the contributions of

individual features to response prediction is more challenging with such models relative to multivariate linear regression models.

**Deep learning.** Deep learning is a form of artificial intelligence that uses neural networks to uncover latent features within the input data that are associated, for example, with treatment response<sup>93</sup> or survival time<sup>94</sup>. Although these models offer more flexibility in learning, the feature transformations often make providing a biological context to the artificial intelligence-based biomarkers that are discovered challenging, hence these agnostic models are often referred to as ‘black boxes’.

The extraction of human-interpretable features from histopathological and/or radiological images can be resource-intensive; therefore, leveraging deep learning to mine patterns from image pixels that reflect the morphological and/or anatomical characteristics of tumours has become an attractive approach. Neural networks have enabled the development of imaging-based predictors of known clinical biomarkers, such as microsatellite instability and PD-L1, from both radiological images and histology slides, although none of these methods has thus far been implemented clinically (Table 3). The density of information provided by digital pathology analysis can also enable accurate predictions of molecular profiles<sup>95–99</sup>, which can act as an intermediary for predicting treatment responses. ENLIGHT-DeepPT, a framework that first predicts mRNA expression from histology slides and subsequently provides transcriptomics-based response predictions, is an example of this approach<sup>100</sup>. Alternatively, neural networks can predict treatment response directly from medical images, yielding imaging-based predictors of response (Table 3). The input features of medical images (typically pixels or voxels) carry less inherent biological meaning than defined units such as genes; therefore, interpreting these image-based deep learning models (often considered ‘black boxes’) is often particularly challenging.

**Foundation models.** Foundation models, which are pretrained on large unlabelled datasets and adaptable to various downstream tasks,



**Fig. 2 | Overview of advances in computational methods and data collection.**

Integrative timeline of notable discoveries and contributions to each of four categories (in order from top to bottom): predictive biomarkers, profiling technologies, machine learning and data advancements (notable releases of large datasets or data collections that can be used for biomarker discovery), all of which can contribute to improved predictions of treatment responses. CCLE, Cancer Cell Line Encyclopedia; CPTAC, Clinical Proteomic Tumor Analysis Consortium; CTRP, Cancer Therapeutics Response Portal; dbGaP,

database of Genotypes and Phenotypes; FIMM, Finnish Institute for Molecular Medicine; GEO, Gene Expression Omnibus; HTS, high-throughput screening; LINCS, Library of Integrated Network-based Cellular Signatures; NCI/DTP, National Cancer Institute Developmental Therapeutics Program; NGS, next-generation sequencing; NIBR PDXE, Novartis Institutes for BioMedical Research Patient-Derived Xenograft Encyclopaedia; NN, neural network; RT-PCR, PCR with reverse transcription; SRA, Sequence Read Archive; TCGA, The Cancer Genome Atlas.

**Table 3 | Examples of monotherapy response predictions and their adherence to the hallmarks of predictive oncology**

Name	Model architecture, prediction and interpretability	Performance and clinical relevance
<b>Linear multivariable omics predictors</b>		
MammaPrint <sup>73</sup> (2002)	A logistic classifier for patient stratification based on gene expression data; interpretable by nature	Predicted 5-year DFS with 89.5% accuracy; prospectively identified patients with low-risk, early-stage breast cancer who can safely avoid chemotherapy (5-year DFS 94.7%) <sup>74,218</sup> ; in clinical use
Oncotype DX <sup>83</sup> (2004)	A Cox regression model for patient stratification for recurrence risk based on gene expression data; interpretable by nature	6.8% of patients with breast cancer stratified as having a low recurrence risk had distant recurrence at 10 years, whereas 30.5% of those deemed to have a high risk of recurrence had disease recurrence at this time point <sup>219</sup> ; in clinical use
PAM50 (ref. 84) (2009)	A Cox regression model for patient stratification for neoadjuvant chemotherapy based on gene expression data; interpretable by nature	94% sensitivity for identifying patients with early-stage breast cancer who do not respond to neoadjuvant chemotherapy (NPV 97%); in clinical use
OncoTreat <sup>220</sup> (2018)	A VIPER-based algorithm for drug prioritization based on protein activity inferred from gene expression data; interpretable by nature	PDX models obtained from patients with GEP-NETs had a TGI of 110–112% when exposed to entinostat (the top prioritized drug) versus 8% in PDXs exposed to belinostat
ENLIGHT <sup>155</sup> (2023)	An exponential parametric survival model for drug response prediction based on gene expression data; interpretable by nature	Predicted response to ICIs when combined with IFN $\gamma$ signature (OR >4 versus other candidate biomarkers) in a retrospective analysis of data from 21 clinical datasets
<b>Non-linear multivariable omics predictors</b>		
P-Net <sup>179</sup> (2021)	A biologically informed neural network for patient stratification based on genomics data; interpretable via a pathway-aware hierarchical network	Correctly classified 73% of primary and 80% of metastatic prostate cancers in an independent dataset
NeST-VNN <sup>180</sup> (2024)	A biologically informed neural network for drug response prediction based on genomics data; interpretable via a knowledge map	Successfully predicted CDK4/CDK6 inhibitor sensitivity in patients with metastatic breast cancer, who had longer OS relative to those predicted to be strongly resistant (HR 0.29)
scFoundation <sup>109</sup> (2024)	A transformer-based foundation model evaluated for drug response prediction using single-cell gene expression data; interpretability not assessed	An scFoundation-based model (DeepCDR) predicted cell line IC <sub>50</sub> with a Pearson's correlation coefficient of >0.93
<b>Imaging-based predictors of clinical biomarkers</b>		
Kather et al. <sup>221</sup> (2019)	A CNN for MSI prediction based on histology slides; interpretability not assessed	Successfully predicted MSI-H status in a retrospective analysis of H&E-stained slides obtained from patients with colorectal cancer (AUC 0.84)
Deeply learned score <sup>222</sup> (2021)	A CNN for prediction of PD-L1 status based on PET/CT images; interpretable via class activation map	Discriminated between images obtained from patients with PD-L1-positive and PD-L1-negative NSCLCs in two retrospective and one prospective cohort (AUC 0.82). Predictions of PFS and OS were indistinguishable from those made using IHC-based assessments of PD-L1
Wagner et al. <sup>223</sup> (2023)	A transformer model for prediction of MSI status, <i>BRAF</i> and/or <i>KRAS</i> alterations based on histology slides; interpretable via patch contribution visualization	Predicted MSI status in patients with an AUC of 0.97 when applied to resection specimens and AUC 0.86–0.92 when applied to biopsy samples obtained from external validation cohorts
<b>Imaging-based predictors of response</b>		
Xu et al. <sup>139</sup> (2019)	A model based on transfer learning of CNNs with RNN for drug response prediction based on CT images; interpretability not assessed	Stratified patients with locally advanced NSCLC into high-risk and low-risk groups with significant differences in OS (HR 6.16; 95% CI 2.17–17.44; <i>P</i> < 0.001).
Ogier du Terrail et al. <sup>184</sup> (2023)	A federated learning framework for multiple-instance learning of a logistic regression model for drug response prediction based on histology slides; interpretable via spatial tile scores	Predicted the responses of patients with TNBC receiving neoadjuvant chemotherapy (AUC 0.66 in global validation)
ENLIGHT-DeepPT <sup>100</sup> (2024)	A framework of a CNN, an AE, an MLP and an exponential parametric survival model for drug response prediction based on histology slides; interpretable via gene expression	Predicted true responders across 5 independent cohorts spanning 4 different cancer types and 6 different treatment approaches with a 39.5% increase in response rate (OR 2.28) among predicted responders versus the baseline rate
<b>Multimodal predictors</b>		
HRDetect <sup>75</sup> (2017)	A LASSO logistic regression for patient stratification based on HRD, genomic alterations and mutational signatures; interpretable by nature	Predicted <i>BRCA1/BRCA2</i> deficiencies in patients (AUC 0.98); also enabled the detection of tumours with somatic or functional <i>BRCA1/BRCA2</i> deficiencies with no detectable mutations
Sun et al. <sup>224</sup> (2018)	An elastic net model for immunotherapy (anti-PD-L1) response prediction based on CE-CT images, location variables and a technical variable; interpretability not assessed	Stratified patients with an objective response from those with PD or SD ( <i>W</i> = 0.13; <i>P</i> = 0.025) and patients with durable (6-month) disease control from those with PD ( <i>W</i> = 0.12; <i>P</i> = 0.013)

**Table 3 (continued) | Examples of monotherapy response predictions and their adherence to the hallmarks of predictive oncology**

Name	Model architecture, prediction and interpretability	Performance and clinical relevance
<b>Multimodal predictors (continued)</b>		
DrugCell <sup>17</sup> (2020)	A neural network for drug response prediction based on genomics data and drug chemical structures; interpretable via a knowledge map	Predicted response to either mTOR or CDK4/CDK6 inhibitors: patients with predicted sensitivity had a median OS of 48.2 months versus 33.6 months for those with predicted insensitivity
Transfer of Cell Line Response Prediction <sup>40</sup> (2021)	Few-shot learning on a neural network for drug response prediction based on genomics and gene expression data; interpretable via molecular markers	Predicted disease control (SD, PR or CR) versus progressive disease (ORs 3–10.5) in PDX models exposed to various therapies
Sammut et al. <sup>128</sup> (2022)	An ensemble of an elastic net logistic regression, SVM and random forest for drug response prediction based on clinical information, digital pathology, and genomics and gene expression data; interpretable via feature importance	Predicted pathological CR following standard-of-care neoadjuvant therapy in an independent cohort of 75 patients with breast cancer (AUC 0.87)
DyAM <sup>30</sup> (2022)	A deep attention-based multiple-instance learning model for immunotherapy response prediction based on clinical information, CT images, digitized PD-L1 IHC and genomics data; interpretability not assessed	Predicted response to anti-PD-(L)1 antibodies in patients with advanced-stage NSCLC with an AUC of 0.8, higher than for TMB (AUC 0.61) and PD-L1 IHC score (AUC 0.73)
LORIS <sup>114</sup> (2024)	A logistic regression model for immunotherapy response prediction based on TMB, treatment history, blood markers, age and cancer type; interpretable by nature	Predicted objective response to immunotherapy across 6 validation datasets comprising patients with various advanced-stage solid tumours (AUCs 0.64–0.79), with 15–68% higher AUCs than for TMB
MUSK <sup>121</sup> (2025)	A transformer-based foundation model for patient stratification and immunotherapy response prediction based on histology slides and clinical/pathology reports; interpretable via attention heat maps	Predicted 5-year relapse following curative-intent surgery (AUC 0.83); predicted response (CR or PR) to ICI (AUC 0.77) with a higher AUC than for PD-L1 (AUC 0.61) across 16 cancer types
Clinical Transformer <sup>123</sup> (2025)	A transformer-based foundation model evaluated on patient stratification based on genomics, gene expression, and proteomics data and TME, laboratory test results, and demographic information; interpretable via feature permutation importance	Predicted survival following anti-PD-(L)1 antibodies and combinations (C-index 0.73), with a higher C-index than for TMB (C-index 0.55); stratified patients (HR 0.29) more effectively than with TMB (HR 0.69)

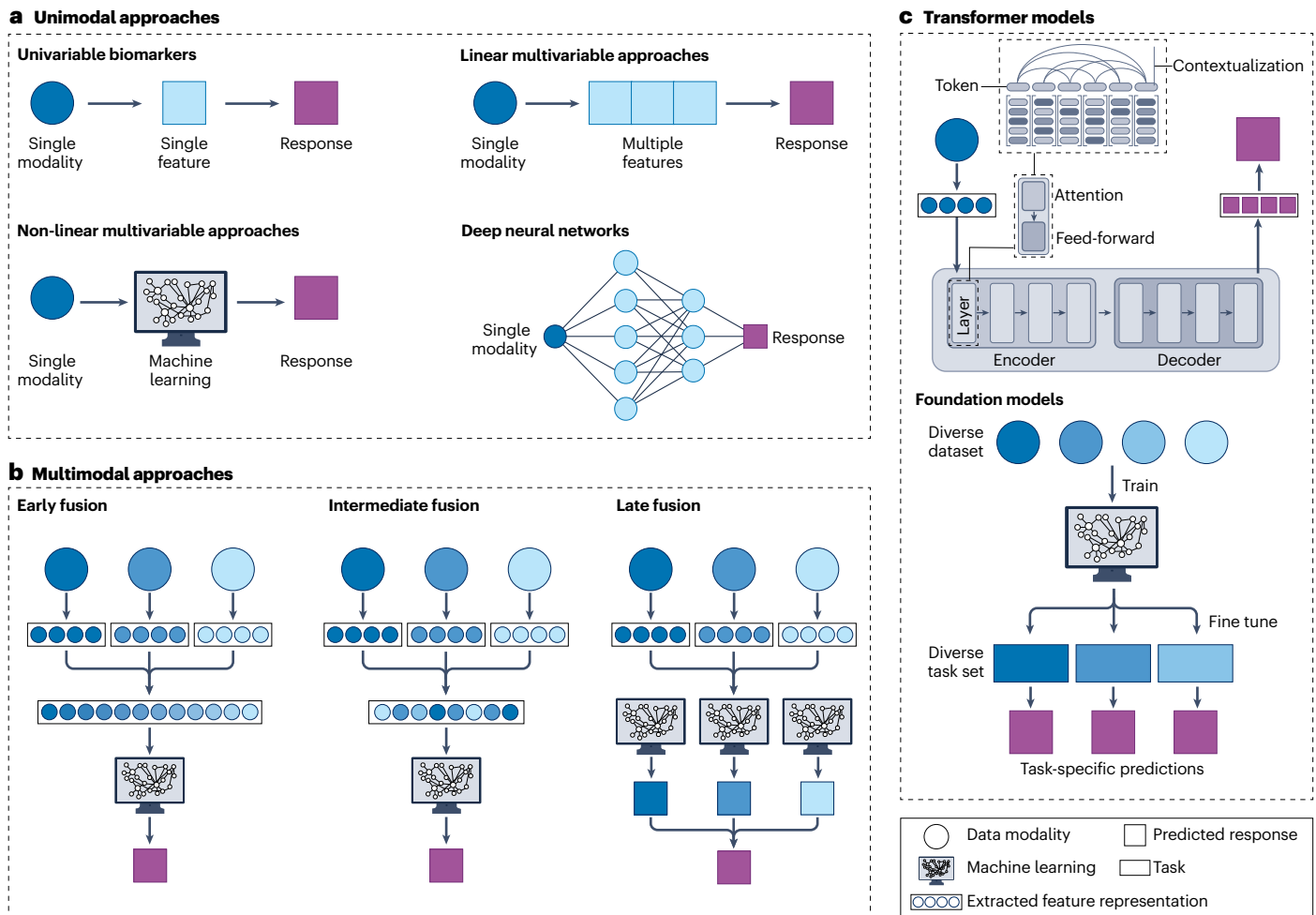
AE, autoencoder; AUC, area under the receiver operating characteristic curve; CE-CT, contrast-enhanced CT; C-index, concordance index; CNN, convolutional neural network; CR, complete response; DFS, disease-free survival; GEP-NET, gastroenteropancreatic neuroendocrine tumour; H&E, haematoxylin and eosin; HR, hazard ratio; HRD, homologous recombination deficiency; IC<sub>50</sub>, half-maximal inhibitory concentration; ICI, immune-checkpoint inhibitor; IHC, immunohistochemistry; MLP, multi-layer perceptron; MSI, microsatellite instability; MSI-H, high MSI; NPV, negative predictive value; NSCLC, non-small-cell lung cancer; OR, odds ratio; OS, overall survival; PD, progressive disease; PDX, patient-derived xenograft; PFS, progression-free survival; PR, partial response; RNN, recurrent neural network; SD, stable disease; SVM, support vector machine; TGI, tumour growth inhibition; TMB, tumour mutation burden; TME, tumour microenvironment; TNBC, triple-negative breast cancer.

have emerged as powerful tools for biomarker discovery. These models are usually transformers, a deep learning architecture that uses self-attention mechanisms to model relationships between all elements in a sequence simultaneously, enabling efficient handling of long-range dependencies<sup>101</sup>, often trained using self-supervised learning, a technique in which models learn latent representations within unannotated datasets and can then be fine-tuned for further training on smaller, labelled datasets to identify predictive biomarkers (Table 2 and Fig. 3). This approach overcomes the need for resource-consuming data annotation and model training for each individual prediction task while also circumventing any limitations in dataset size, diversity and batch effects across specific data collection centres. Several unimodal biomedical foundation models have been developed, including GatorTron<sup>102</sup> for EMRs; FMCIB<sup>103</sup> for radiomics; UNI<sup>104</sup>, Prov-GigaPath<sup>105</sup> and CHIEF<sup>106</sup> for pathology; scGPT<sup>107</sup>, Geneformer<sup>108</sup> and scFoundation<sup>109</sup> for single-cell transcriptomics; and scGPT-spatial<sup>110</sup> for spatial transcriptomics. scFoundation has been further fine-tuned on bulk-level gene expression data from preclinical pharmacogenomics datasets to accurately predict drug responses in cell lines, demonstrating the applicability of foundation models for response prediction<sup>109</sup>. With increasing data availability and rapid developments in multimodal integration, the widespread use of foundation models capable of

leveraging clinical, imaging and omics data for biomarker discovery is imminent<sup>111</sup>.

### Multimodal approaches

Restricting biomarker discovery to individual modalities or feature types might yield only limited representations of cancer dynamics and tumour heterogeneity. For example, TMB can guide the use of ICIs in certain scenarios, although multimodal approaches have been shown to improve response stratification<sup>79,112,113</sup> with an example provided by the six-feature LORIS signature that integrates TMB with PD-L1 expression as well as more traditional clinical characteristics<sup>114</sup>. Several other multimodal approaches indicate similar improvements in performance beyond that of individual modalities alone. An example is provided by a deep learning model trained on radiomics, digitized PD-L1 immunohistochemistry slides and genomics data, including TMB, which was found to predict response to anti-PD-(L)1 antibodies with an area under the curve (AUC) of 0.8, outperforming LASSO logistic regression models trained on the individual component features, including TMB (AUC 0.61), PD-L1 immunohistochemistry score (AUC 0.73), CT features (AUC 0.28–0.65) and genomic alterations (AUC 0.61)<sup>30</sup>. Given the advantages of comprehensive multimodal predictors over single-modality approaches, growing interest exists in the application of multimodal models for predicting treatment response (Table 3).



**Fig. 3 | Overview of computational methods for biomarker discovery.** Different computational approaches for biomarker discovery. **a**, Computational methods can identify single features that predict response (univariable biomarkers). Alternatively, multiple features can be combined linearly (signatures) or non-linearly (machine learning or deep learning) to predict response. **b**, Approaches involving multimodal data fusion to discover multimodal predictors comprise three main categories. Feature vectors of different modalities can be directly concatenated for model input (early fusion), transformed to obtain new features for model input (intermediate fusion), or analysed using unimodal models to

obtain single-modality-based predictions that can then be integrated into a single final prediction (late fusion). **c**, Transformer models: overview of transformer architectures. Model inputs (typically text or numerical vectors) are passed into encoder layers, where they are parsed into units called tokens, which are contextualized to surrounding tokens by attention mechanisms. Foundation models include deep learning models, such as transformers, which are pretrained on large datasets to learn general contextualizations of data before being re-trained (fine-tuned) on smaller datasets for more specific prediction tasks, enabling the model to be applied to a diverse task set rather than a single task.

Strategies for fusing multimodal data have traditionally been classified as using early (raw feature level), intermediate (within learned latent spaces) or late (single-model outputs) fusion approaches, denoting the point within the computational pipeline at which data representations from different modalities are combined<sup>115</sup> (Fig. 3). Early fusion strategies combine data at the input level such that the model evaluates all features simultaneously. Late fusion architectures train separate models for each modality before aggregating predictions with consensus voting, averaging and Bayes-based rules, among other methods<sup>116,117</sup>. Intermediate fusion strategies first learn representations for each modality, which are then integrated before providing a final joint prediction. Intermediate fusion might lead to increased model complexity, although this approach also provides flexibility in capturing feature interactions across model architectures. Several

techniques for multimodal intermediate fusion exist, such as deploying individual encoders for each data modality to obtain a merged feature space<sup>118</sup>, or creating individual graphical representations for each modality that can then be fused to create a unified multimodal graphical representation<sup>119</sup>. Over the past few years, transformer-based architectures, along with contrastive learning, have gained popularity as methods of capturing the complex relationships between different feature types<sup>120</sup>. A biomedical application of these models involves the development of vision–language models that integrate patient imaging and EMR data using contrastive self-supervised learning, a technique that identifies similarities between images and text. An example of this application is provided by MUSK, a transformer-based vision–language foundation model pretrained using image–text pairs via contrastive learning for the prediction of disease relapse in patients

with melanoma as well as response to ICIs in patients with NSCLC or gastro-oesophageal cancer based on clinical reports and slides stained with haematoxylin and eosin<sup>121</sup>. With the rapid evolution of deep learning and the increasing availability of sequencing data, we anticipate that the next generation of multimodal foundation models will leverage multi-omics profiling. These transformer architectures will be able to generate tokens based on molecular units, such as nucleotides, genes and proteins, to enable cross-modality attention operations<sup>122</sup>. An example of this approach is provided by the Clinical Transformer framework, which was pretrained on multi-omics, clinical and patient-demographic data from large datasets before being transferred to smaller clinical datasets to predict response to ICIs<sup>123</sup>. Such deep learning architectures using multi-headed attention for multimodal fusion have the potential to facilitate clinical, imaging and omics data integration<sup>116,124</sup>.

In designing multimodal models, different integration and fusion techniques should be assessed to identify the best approach for the data types being used. Furthermore, the incorporation of multiple modalities requires considerations regarding data availability, given that routine clinical diagnosis typically does not encompass the collection of certain types of omics data. As a result, the lack of sufficient input data from standard clinical workflows might reduce the robustness and clinical applicability of discovered predictive models when applied to real-world patient populations. Predictive models trained on multimodal data that include sources not routinely collected in clinical practice often have limited real-world applicability as missing modalities can compromise predictive accuracy and hinder clinical integration.

## Considerations for feature and model selection

Selecting the most suitable modelling approach and combination of feature types for predicting treatment outcome is not a straightforward task. With increasing model complexity and the growing number of data modalities, coupled with the small sizes of many datasets, predictive accuracy might be improved, albeit at the cost of potential overfitting and an increased risk of false discoveries. This dilemma, known as the bias–variance trade-off, is a predominant reason as to why multivariable and multimodal models have largely failed to be implemented clinically<sup>125</sup>. In an attempt to mitigate these challenges, feature selection techniques are widely used to reduce dimensionality and improve the signal-to-noise ratio prior to model training<sup>87,126</sup>. Identifying and focusing on the most informative features can improve both the statistical robustness and interpretability of the discovered biomarkers, ultimately supporting their clinical implementation. Models such as regularized linear regression and forest-based approaches implement additional feature selection by either shrinking or removing less informative features within their architectures. Other techniques, such as regularization, dropout, pruning and cross-validation, are also commonly used to mitigate overfitting. Another standard model-development practice involves benchmarking performance against that of different models of various complexities and combinations of feature modalities to optimize both model performance and generalizability for the prediction task at hand, often following the principle of Occam's razor<sup>125</sup>.

Multiple models can also be integrated into an ensemble such that their individual predictions are combined to produce a single final prediction. This approach has been shown to reduce the extent of model-specific biases and improve overall predictive performance<sup>89,127</sup>. For example, an ensemble leveraging clinical, genomic, transcriptomic and digital pathology features was developed to predict the responses to neoadjuvant therapy of patients with breast cancer and achieved a

superior level of predictive accuracy over that of other combinations of feature inputs and models<sup>128</sup>. Alternatively, running multiple iterations of the same model can generate a distribution of predictive signatures, enabling the identification of stable and robust features. Beyond variability attributed to individual models, the choice of a single response variable might also yield a less comprehensive prediction. To this end, multitask learning – a technique based on the belief that predictive accuracy can be improved by learning multiple related prediction tasks<sup>129,130</sup> – can be leveraged to improve model performance by using multiple response variables. For example, a hypothetical model trained to predict TMB, PD-L1 status, objective response (defined by RECIST), overall survival and risk of toxicity for patients receiving immunotherapy can not only provide a more representative prediction of treatment outcomes but also improve on the accuracy of each individual prediction task by leveraging complementary information derived from learning other tasks. Models trained using these practices might provide a more generalized representation of the biological signals being measured.

For models trained using preclinical data, model selection should also be cognisant of the translational relevance of such data to patients. The wealth of preclinical models available has fuelled the development of complex predictive models such as Bayesian multitask and multisource learning methods for the prediction of responses to BCL-2 inhibitors *ex vivo* in samples obtained from patients with acute myeloid leukaemia<sup>131,132</sup>. However, clinical response predictions have historically leveraged simpler methods. For example, Oncotype DX<sup>83</sup>, a clinically validated 21-gene assay that predicts benefit from chemotherapy in patients with breast cancer, was discovered using a multivariable Cox regression model trained on prospectively obtained data. Major limitations to the development of these more complex clinical predictors include the lack of availability of clinical data (Box 1) as well as the sometimes-limited alignment between preclinical and clinical data, which has challenged the successful translation of effective preclinical models to patient data. Furthermore, patient responses are measured using clinical end points and, therefore, predictive models trained using preclinical samples cannot be directly applied to patient data owing to the use of different measurements of efficacy. Few-shot learning<sup>40</sup> and transfer learning<sup>39</sup> approaches have the potential to improve the accuracy of clinical predictions on tasks with limited data by leveraging knowledge from models pretrained on large datasets and then fine-tuning or adapting them to new contexts, enabling accurate predictions despite minimal examples. These learning approaches are expected to greatly improve the predictive power of biomarker discovery efforts and provide a more comprehensive understanding of both the efficacy of various therapies and mechanisms of resistance<sup>40,133</sup>. Such approaches provide opportunities to maximize the utility of preclinical data for model training while increasing the likelihood of providing clinically relevant response predictions.

Another major challenge with predictive models is the capacity to learn from new data without succumbing to catastrophic forgetting, a phenomenon in which models 'forget' previous information rather than retaining and integrating with the new information<sup>134</sup>. The ability of artificial intelligence and machine learning models to continuously learn data while retaining previous knowledge and predictive abilities, known as continual learning, can enable treatment response predictors to effectively take advantage of new data as it emerges within clinical settings without having to retrain from scratch. Several continual learning techniques, such as elastic weight consolidation<sup>134</sup>, among others<sup>135–138</sup>, are gaining popularity across several machine learning-relevant domains, and the autodidactic learning abilities of neural networks would further support the implementation of such

continual learning approaches. However, applications of these techniques as methods of predictive biomarker discovery can be limited by regulatory hurdles as models tend to be 'locked' (and therefore cannot be changed over time) to prevent the introduction of bias or deviations from the initial training data. In a similar vein, predictive models should be able to leverage longitudinal data to provide updated treatment predictions based on the disease trajectories of each patient, including evolution of disease biology under the selective pressures created by treatment. For example, a neural network architecture trained on baseline CT images and then post-treatment scans obtained at three follow-up time points to predict response to chemoradiotherapy in patients with NSCLC showed that model performance improves with the addition of data from each follow-up scan<sup>139</sup>. By incorporating continual learning techniques, predictive models can remain dynamic tools that are robust and adaptive to new clinical insights and changes along a patient's treatment journey.

## Combination therapy biomarkers

Anticancer combination regimens are often more effective than monotherapies and might reduce the likelihood of resistance to, as well as potentially reducing the doses of, each agent that needs to be administered for therapeutic activity<sup>140,141</sup>. Various combinations involving multiple chemotherapies and/or chemotherapy plus an ICI are now the first-line standard-of-care approach, for example, FOLFIRINOX or gemcitabine–nab-paclitaxel as adjuvant therapy for patients with pancreatic cancer<sup>142,143</sup>, pembrolizumab plus chemotherapy for early-stage triple-negative breast cancer<sup>144</sup>, or chemoradiotherapy followed by consolidation durvalumab for unresectable stage III NSCLC<sup>145,146</sup>. However, further optimization of outcomes as well as opportunities for personalization are limited because most clinically used combinations were identified on the basis of empirical testing, resulting in a lack of biomarkers capable of predicting which patients are most likely to benefit from a specific combination regimen<sup>147</sup>.

Similar artificial intelligence and/or machine learning approaches can be used to predict benefit from both monotherapy and combination therapy (for example, classifying responders versus non-responders following similar modelling strategies), although the mechanisms

of action underlying responses to combination therapy are typically more heterogeneous than those of monotherapies, resulting in the need for more complex prediction approaches. Several unique computational challenges related to combination therapy predictions also exist, including the much larger range of potential drug combinations to be considered for each patient and the lack of clinical response data for many combinations not yet tested in clinical trials or translational studies for artificial intelligence and/or machine learning model training and testing. Similar to predictions related to benefit from monotherapy, outcomes to be predicted for combination regimens include therapeutic efficacy and toxicities; however, in many studies, synergistic activity of the regimen is also considered an important outcome owing to the potential to enable reductions in the doses of one or more drugs included in the regimen and thus improve tolerability. Indeed, when developing combination regimens, the therapeutic index, meaning the balance between therapeutic effect and toxicities, is an especially important factor to consider. However, dosage or toxicity predictions, neither from traditional preclinical nor *in silico* models, are able to accurately inform on the optimal therapeutic doses or risks of adverse effects of specific regimens prior to testing in patients.

## Recommender systems for patient matching

Current biomarker-based clinical trials, such as I-PREDICT<sup>148</sup>, use large panels of established DNA-based biomarkers to identify and test novel combination regimens in patients with cancer. These studies have demonstrated that targeting a larger range of patient-specific alterations than those included in the often limited gene panels used to match patients with specific alterations to the most appropriate monotherapies might improve disease control rates without increasing the risk of adverse effects<sup>149</sup>. Similarly, the NCI-ComboMATCH study aims to overcome resistance to monotherapies and leverage novel mechanisms of synergy to improve efficacy using genomics-based matching of patients to combination therapies<sup>150</sup>. Beyond the use of such predefined gene and drug panels, more systematic matching-based computational recommendation systems based on each patient's molecular signatures (Box 2) should leverage multi-omics prediction approaches to identify the most effective, synergistic and/or least toxic combination regimens

## Box 2 | Prediction tasks for computational treatment response optimization

Non-mutually exclusive prediction tasks addressed by computational methods can support various types of clinical application such as matching patients to a specific treatment regimen or identifying candidate treatments for a given patient. Here, we provide several examples of such systems and their roles in response optimization.

**Recommender systems for patient–treatment matching:** these systems provide personalized suggestions of the most effective drugs and/or treatment combinations from a range of pre-selected options (either based on binary matching between patients and treatments, or ranking of possible treatment options for each patient). Recommender systems have been implemented as matching algorithms in biomarker-based trials (such as NCI-ComboMATCH, which uses the MATCHBox algorithm<sup>150</sup>).

**Treatment-tailored predictive models:** a single model trained or fine-tuned using the available data on the activity of the particular

treatment option (monotherapy or combination) and subsequently applied to data from each patient to obtain a quantitative efficacy score reflecting the predicted treatment response (for example, to CDK4/CDK6 inhibitors in patients with metastatic breast cancer<sup>17</sup>). Such models can guide the selection of patient subpopulations in histology-specific biomarker-based clinical trials.

**Patient-tailored predictive models:** a separate model trained or fine-tuned for each patient using patient-specific data to predict probable responses to multiple treatment options; these models provide an efficacy score enabling treatment selection from a prespecified range of treatments (potentially all approved drug combinations). Such models can be used to support the general tailoring of treatment approaches for precision oncology applications. Examples include ENLIGHT<sup>155</sup> for monotherapy and scTherapy<sup>166</sup> for combination prediction.

that optimize therapeutic benefit<sup>151</sup>. An example of such an application is provided by MCWI-PREDICT, an ongoing prospective trial leveraging multi-omics profiling for treatment selection purposes<sup>152</sup>.

## Efficacy, synergy and toxicity

The number of possible drug combinations vastly exceeds what can be tested clinically and, therefore, artificial intelligence and/or machine learning models might provide a useful method of narrowing down the number of candidates and prioritizing the most effective drug combinations to be tested in preclinical models. The search objective is to identify combinations that result in either synergistic effects or have additive efficacy. Owing to the high levels of molecular and phenotypic heterogeneity of cancer cells, clinically relevant synergistic activity is extremely rare and is often highly dependent on the cell context<sup>147</sup>. This rarity poses challenges to the identification of biomarkers associated with responsiveness to combination regimens and requires large panels of cell lineages that capture various cellular contexts and drug treatment phenotypes. Large-scale and clinically representative datasets are therefore needed to identify the most effective combination for each individual cancer, instead of broadly active combinations.

## Combination treatment-tailored prediction models

Owing to the limited amount of data available on the outcomes of patients receiving combination regimens relative to preclinical data, most computational prediction methods for identifying features associated with a response to particular regimens have been developed using data from large-scale cell line pharmacogenomics resources. The current computational methods predominantly focus on the prediction and/or classification of synergistic versus additive effects of specific regimens (Table 4), some of which have been benchmarked<sup>153</sup> and described in detail elsewhere<sup>154</sup>. For example, the deep learning model DrugCell was trained on genomic and phenotypic response data from large collections of cancer cell lines<sup>17</sup>. This model maps states in cellular subsystems induced by tumour genotypes that are most predictive of treatment responses, thereby also revealing the biological mechanisms underlying these responses. DrugCell predictions were not only accurate in cell lines and PDX models but were also predictive of the responses of patients with metastatic breast cancer to mTOR and CDK4/CDK6 inhibitors. This study provides an example of a treatment-tailored prediction method with predictive relevance to a specific cancer type and treatment approach (Box 2).

## Patient-tailored drug combination predictions

An alternative treatment-matching approach involves prediction of the response of each patient to a variety of regimens, including chemotherapies, targeted therapies, radiotherapy and immunotherapies to identify the most effective approach for each particular patient (Box 2). For example, ENLIGHT implements a transcriptomics-based survival modelling approach that leverages a wide range of cancer cell line data to identify clinically relevant genetic synthetic lethal interactions. Responses to multiple therapy options are predicted even in the absence of clinical data on treatment responses<sup>155</sup>. ENLIGHT supports two types of translational application: (1) personalized treatment selection that enables selection of the most effective treatments for each patient; and (2) clinical trial design that aims to select the most likely responders in a given cohort. Retrospective analyses suggest that applying ENLIGHT for patient stratification purposes can increase ORRs in clinical trials testing immunotherapies and other monoclonal antibodies by excluding predicted non-responders, in some cases resulting in

>90% of the ORR attainable with an optimal exclusion strategy (in which every non-responder is excluded prior to treatment). In addition to standard artificial intelligence and/or machine learning models such as ENLIGHT, deep transfer learning models that are first pretrained on large-scale cell line data and then fine-tuned with patient data could further improve the prediction of patient-specific drug combination responses *ex vivo*<sup>156</sup>.

## Single-cell profiling for cancer-selective combinations

Compared with bulk molecular data, single-cell profiling enables the identification of cell-type-specific biomarkers predictive of treatment responses in various cancer cell and non-malignant populations. For example, emerging single-cell profiling atlases offer information on intratumoural cellular heterogeneity that can be leveraged to identify combinations that selectively target drug-resistant cancer cells in patients with relapsed disease<sup>157</sup>. Treatment approaches informed by single-cell profiling might even enable resistance to first-line therapies to be avoided by co-inhibiting multiple cancer cell subpopulations, including those that drive disease progression and/or resistance. Single-cell data also enables investigators to design regimens that leverage the effects of two or more drugs on specific cell populations for improved patient outcomes<sup>158</sup>. Several research groups have developed computational models that predict single-cell transcriptomics or other *in vitro* or *ex vivo* cellular responses to targeted therapies as well as identifying genetic perturbations in cell lines and patient-derived samples<sup>159–163</sup>. Similarly, single-cell RNA sequencing (scRNA-seq)-based profiling has been used together with *ex vivo* drug response profiling in the context of distributed gradient-boosted decision trees to identify synergistic and minimally toxic combination regimens for individual patients with acute myeloid leukaemia, thereby increasing their likelihood for clinical translation<sup>157,164</sup>.

As an example of a method validated using clinical data, PERCEPTION pretrains a regularized regression model for a given treatment regimen by leveraging bulk sequencing and scRNA-seq cell line profiling data to predict clonal responses. This model successfully predicted responses to targeted therapies in patient-derived primary cells as well as in data from two clinical trials involving patients with multiple myeloma receiving daratumumab plus carfilzomib, lenalidomide, and dexamethasone or patients with breast cancer receiving endocrine therapy plus a CDK4/CDK6 inhibitor, in which it distinguished responders from non-responders with AUCs of >0.83 and >0.78, respectively<sup>165</sup>. However, training such a treatment-tailored predictive model for a larger number of potential combinations is impractical. By contrast, a patient-tailored treatment-prediction approach, named scTherapy, enables the identification of specific targeted combinations for an individual patient based solely on scRNA-seq data obtained from the analysis of tumour samples<sup>166</sup>. To enable rapid translational application, a gradient-boosting model was pretrained to learn differences in drug responses across cellular populations using cell line data. When applied to a patient sample, the model generates a ranked list of the most effective multi-targeting options with effective doses that are predicted to have a selective synergistic activity against key cancer cell clones in each sample. Experimental validation in patient-derived *ex vivo* models confirms that 96% of the predicted effective treatments demonstrated selective efficacy or synergy, and 83% were minimally toxic to non-malignant cells.

The adoption of single-cell profiling in clinical settings has thus far been limited, although these methods have started to emerge as routine profiling technologies in research settings and for patient

**Table 4 | Examples of drug combination response prediction models**

Name	Model architecture and prediction	Performance
DeepSynergy <sup>225</sup> (2018)	A DNN that predicts Loewe additivity values based on gene expression data and drug connectivity fingerprint	7.2% more effective at predicting novel combinations than other machine learning methods within the context of explored drugs and cell lines in the Merck 2016 dataset; AUC 0.9 when applied to the classification of novel drug combinations
Xia et al. <sup>226</sup> (2018)	A DNN that computes a modified ComboScore based on gene expression, miRNA expression, protein abundance, drug path fingerprint, extended drug connectivity fingerprint and drug categorical descriptors	Predicted drug combinations from the NCI-ALMANAC dataset (PCC 0.97 and Spearman correlation coefficient 0.96)
AuDNNsynergy <sup>227</sup> (2021)	A DNN plus AE that predicts Loewe additivity values based on genomics and gene expression data and extended drug connectivity fingerprint	Predicted drug synergy from the Merck 2016 dataset (rank correlation coefficient 0.73 and AUC 0.91)
Kim et al. <sup>228</sup> (2021)	A transformer plus DNN that predicts Loewe additivity values based on gene expression data, MACCS fingerprints and SMILES strings	Predicted drug synergy from the Merck 2016 dataset (AUC 0.96 for binary classification task and MSE 174.3 for regression task)
GraphSynergy <sup>229</sup> (2021)	A GNN that predicts Loewe additivity values based on PPI, cancer cell line–protein associations	Predicted drug synergy from the DrugCombDB and Oncology-Screen datasets (accuracy 0.75), with 11–11.9% improved performance versus DeepSynergy
TransSynergy <sup>230</sup> (2021)	A DNN plus transformer that predicts Loewe additivity values based on gene expression data and multi-hot encoded SMILES strings	Predicted drug synergy from the Merck 2016 dataset (Spearman correlation coefficient 0.73, PCC 0.75, AUC 0.91)
MatchMaker <sup>231</sup> (2022)	A DNN that predicts Loewe additivity values based on gene expression data and ChemoPy fingerprint	Predicted drug synergy from the DrugComb dataset (Spearman correlation coefficient 0.74, PCC 0.79 and AUC 0.97)
PRODeepSyn <sup>232</sup> (2022)	A GNN that predicts Loewe additivity values based on PPI, gene expression and genomics data, Morgan fingerprint, and drug categorical descriptors	Predicted drug synergy from the Merck 2016 dataset (PCC 0.75, AUC 0.9)
DTSyn <sup>233</sup> (2022)	A transformer plus GNN that predicts Loewe additivity values based on gene expression data and DeepChem-processed SMILES strings	Predicted drug synergy from the Merck 2016 dataset (AUC 0.73–0.82)
SynPathy <sup>234</sup> (2022)	A DNN that predicts Loewe additivity values based on pathway enrichment analysis of genomics data, Morgan fingerprint and drug target molecular pathway enrichment analysis	Predicted drug synergy from the Merck 2016 dataset (MSE 64.7)
CCSynergy <sup>235</sup> (2023)	A DNN that predicts Loewe additivity based on gene expression data, signalling pathway activity, transcription factor activity, CRISPR-based gene essentiality, DepMap signalling pathway dependencies, Morgan fingerprint, drug 3D fingerprint, Murcko's scaffold, and drug medical and physicochemical properties	CCSynergy III, IV and V predicted drug synergy from the Merck 2016 dataset (PCC >0.81); CCSynergy III and V predicted drug synergy from the Sanger dataset (AUC >0.84)
DFFNDDS <sup>236</sup> (2023)	A DNN plus transformer that predicts Loewe additivity values based on gene expression data and SMILES strings	Predicted drug synergy from DrugCombDB (AUC 0.92) and from DrugComb dataset (AUC 0.85)
MARSY <sup>237</sup> (2023)	A GNN plus transformer that predicts ZIP scores based on gene expression data	Predicted drug synergy from the DrugComb dataset (PCC 0.89, Spearman correlation coefficient 0.78)
MGAE-DC <sup>238</sup> (2023)	A DNN plus GNN plus AE that predicts Loewe additivity values, Bliss values, ZIP scores, and HSA based on gene expression data and drug molecular fingerprints	Predicted, from the Merck 2016 dataset, Loewe (PCC 0.83), Bliss (PCC 0.84), ZIP (PCC 0.85) and HSA (PCC 0.83)
SNRMPACDC <sup>239</sup> (2023)	A DNN plus CNN that predicts Loewe additivity values based on gene expression and genomics data, ECFP_6, and drug physicochemical and toxicophore characteristics	Predicted drug synergy from the Merck 2016 dataset (PCC 0.75 and AUC 0.91)

AE, autoencoder; AUC, area under the receiver operating characteristic curve; CNN, convolutional neural network; DNN, deep neural network; ECFP\_6, counts of extended drug connectivity fingerprints with a radius of 6; GNN, graph neural network; HSA, highest single agent; MACCS, Molecular Access Systems; miRNA, microRNA; MSE, mean squared error; PCC, Pearson correlation coefficient; PPI, protein–protein interaction; SMILES, Simplified Molecular Input Line Entry System; ZIP, zero interaction potency.

stratification in prospective trials<sup>54,167–169</sup>. Newer methods, such as PromethION<sup>170</sup>, a nanopore sequencing platform, can reduce the costs of single-cell sequencing to <US\$1,000 per sample. Single-cell analysis technologies provide deep insights into drug activity and distribution in non-malignant and malignant tissues, although the large volumes of data generated pose new challenges for both computational modelling and biomarker discovery. As single-cell technologies and computational methods advance and become more robust, we anticipate novel applications, including models designed to predict treatment responses to specific therapies or regimens

and to identify cell-type-specific biomarkers, not only for combinations of small molecules but also when combined with ICIs or other immunotherapies<sup>171</sup>. Similarly, new techniques and devices, such as implantable microdevices capable of delivering multiple drugs into different areas of the same tumour, are expected to provide novel insights into responses to combination regimens, including those involving cancer immunotherapies<sup>172</sup>. Key challenges to solve include guaranteeing sufficiently fast turnaround times to guide treatment decision-making as well as the need for appropriate prospective validation across various indications and regimens.

## Open computational challenges

### Adhering to the hallmarks of predictive oncology

We recently proposed, as part of a larger group of authors, seven hallmarks to address the challenges that impede the clinical translation of predictive oncology models<sup>125</sup>. One such challenge is the lack of transparency and trust in the decisions made using classical artificial intelligence, machine learning and/or deep learning models, which can be overcome using secondary methods to explain model predictions<sup>173–178</sup>. SHAP (Shapley additive explanations), a commonly used feature explainability framework for classical machine learning models, computes the contribution of each input feature to a model's final prediction<sup>173</sup>. Neural networks trained on images can leverage methods such as class activation mapping, which identifies regions of the image to localize features contributing to model predictions, thus supporting human interpretation of these otherwise 'black box' models<sup>175</sup>. Interpretability of deep learning models can also be improved by using model architectures reflecting known biological networks and pathways that enable identification of the key biological processes contributing to specific therapeutic predictions. Examples of these 'biologically informed' deep learning models include DrugCell<sup>17</sup>, P-NET<sup>179</sup> and NeST-VNN<sup>180</sup>.

The lack of model validity remains a major challenge and attempts to address this issue can be hindered by limited access to relevant clinical data, which can reduce both the size and quality of training datasets as well as the availability of relevant cohorts for model validation (Box 1). The lack of standardized clinical annotation, omics profiling and imaging data complicates the use of multiple independent cohorts and data modalities for this purpose. For example, different response evaluations across cohorts, such as variations in the landmark readout times, ORRs (via RECIST), or different response assessment intervals, introduces additional considerations relating to model design and the interpretation of model predictions. Furthermore, as the number of assays required for biomarker assessment increases, delays in obtaining complete molecular profiles may result in left truncation bias, whereby patients with early events (such as rapid disease progression or death before completion of biomarker characterization) are systematically excluded from downstream analyses, thereby limiting the applicability of the model to real-world patients. Collectively, these factors underscore the importance of developing models that not only account for such biases but also align with the principles of IDEA (inclusion, diversity, equity and accessibility) to ensure broad representation across patient demographics.

Open science is an ongoing effort to make scientific research transparent, reproducible and reusable<sup>181</sup>. The sharing of all relevant materials, including but not limited to data, code, documentation and model architectures, should follow the FAIR principles (findability, accessibility, interoperability and reusability)<sup>182</sup>. Federated learning can overcome challenges related to data sharing by enabling the transfer of computational models between participating institutions for local training on patient data without the need to share sensitive data across institutions. This approach enables improvements in predictive accuracy and generalizability beyond models trained on single-institution datasets<sup>183</sup> as demonstrated by the improved chemotherapy response predictions provided by the application of a federated learning infrastructure to the prediction of responses to neoadjuvant chemotherapy in patients with triple-negative breast cancer<sup>184</sup>. Responsible sharing of the necessary components for biomarker discovery, including but not limited to data, model architecture, parameters and documentation, can encourage multicentre collaborations, improve data availability,

and enable specific predictive models to be scrutinized, improved and reused. Several federated artificial intelligence learning networks have been established globally, including the Cancer Artificial Intelligence Alliance, Mayo Clinic Platform, Marathon of Hope Cancer Centres Network<sup>185</sup> and DigiONE<sup>186</sup>. However, widespread deployment of federated learning is often hindered by privacy and security concerns, challenges related to overhead communication, and heterogeneity in both data and hardware across participating institutions<sup>187</sup>. Substantial challenges remain, although the field continues to progress towards open science, and we foresee that this trajectory will support the increased adoption and performance of predictive models in clinical settings.

These hallmarks of predictive oncology can act as guidelines to support refinement and adoption of predictive oncology models<sup>125</sup>. We leveraged several of these hallmarks here to evaluate a selection of models designed to predict benefit from monotherapy (Table 3). Through this evaluation, we demonstrate patterns and challenges in adherence to these hallmarks.

### Addressing the complexity of model development

**Feature selection.** With the increasing availability of data modalities and modelling techniques comes the challenge of identifying the most relevant features (including biological features such as mutations in specific genes as well as machine learning-defined features such as those extracted from neural networks) predictive of clinical outcomes. Selecting a subset of the most informative features can support clinical translation as minimal biomarker panels are easier to implement in clinical environments, both in terms of costs and resources needed. Features should reflect patient-specific tumour characteristics and thus ensure model generalizability; therefore, models trained using preclinical data should include additional measures to minimize inconsistencies in the feature spaces between preclinical models and patient features. For emerging modalities, such as spatial multi-omics, establishing the optimal approaches for deriving quantitative and clinically actionable features for biomarker discovery remains an area of active development.

**Prediction confidence scoring.** Prediction confidence scores provide estimates of certainty for individual predictions from a given model, and considering only predictions that meet a certain confidence threshold has been demonstrated to improve the accuracy of artificial intelligence and/or machine learning models<sup>188,189</sup>. This additional layer of transparency enables clinicians to identify reliable, high-confidence predictions to guide decision-making. Investigating predictions labelled with low confidence can also identify populations or scenarios in which model performance is consistently poor, highlighting opportunities for model improvement. Nonetheless, most of the available treatment response prediction models do not provide prediction confidence scores. Furthermore, the lack of standardization of these values poses challenges relating to interpretation and model performance comparisons.

**Causal machine learning.** Traditional artificial intelligence and machine learning algorithms measure the associations between clinical features and treatment response to predict patient outcomes, whereas causal artificial intelligence and/or machine learning aim to identify the causal relationships between these variables. In other words, this approach estimates treatment effect – the measured improvement in patient outcomes directly caused by a given treatment relative to an alternative treatment regimen – based on individual

or subpopulation-specific traits to support personalized treatment decision-making<sup>190</sup>. Investigating causality yields more generalizable predictors and helps navigate complex scenarios by increasing the interpretability of model-derived predictions. However, implementing this approach is unlikely to be a trivial endeavour owing to the abundance of potential confounding factors coupled with incomplete patient profiling and the limited availability of longitudinal data.

## Challenges in representing spatial and temporal tumour dynamics

Tumour heterogeneity and clonal evolution are key factors to consider when attempting to model and/or predict drug resistance; however, these spatial and temporal dimensions are not well captured by existing clinical biomarkers. Single-cell and spatial multi-omics are emerging technologies that enable higher-resolution, spatially informed tumour investigations; however, the high costs of such assays often limit data availability and scalability beyond tissue samples obtained at single time points (versus serial sampling for temporal assessments and/or collection of multiple tissue samples to capture intratumoural heterogeneity). The temporal dimension of cancer heterogeneity can be captured using serial liquid biopsy – a promising approach for extracting longitudinal biomarkers from circulating tumour-derived materials for monitoring therapeutic response and estimating time to treatment failure<sup>28,191</sup>. Although the low costs and minimal invasiveness make serial analysis of liquid biopsy samples ideal for both *ex vivo* drug sensitivity screening and other clinical applications, these samples provide minimal spatial information. Furthermore, inconsistencies in data extraction methods and collection time points, along with small sample sizes, can make computational endeavours for biomarker discovery challenging<sup>192</sup>. With large ongoing initiatives to generate data across spatial and temporal dimensions<sup>193</sup>, we expect the adoption of biomarker applications designed to consider temporal and spatial tumour dynamics to provide more timely and personalized treatment decisions.

## Translational challenges Pathways for clinical adoption

We anticipate that next-generation predictive biomarkers derived from computational models will incorporate multimodal and multi-omics profiling; however, these technologies remain largely confined to research settings and clinical deployment is currently limited by high costs as well as accessibility and logistical barriers. Current clinical practice largely relies on more pragmatic univariable biomarkers, such as targeted sequencing panels, which offer a balance between breadth of information and feasibility<sup>194,195</sup>. However, these minimalist panels often fail to capture the full extent of biological variability. Deep learning approaches incorporating all available clinical and pre-clinical data are clearly able to provide more accurate predictions than simpler artificial intelligence or classical machine learning models trained on single features or modalities. Nonetheless, data scarcity is a central barrier to the clinical adoption of such approaches (Box 1). Here, generative artificial intelligence has demonstrated utility in synthetically generating preclinical molecular and drug response data<sup>196</sup> and in projecting clinical trial responses onto real-world patients<sup>197</sup>, which have the potential to address the challenges related to limited data availability. Even so, fidelity and the intrinsic biases of synthetically generated data need to be carefully scrutinized. Real-world clinical oncology is constrained by the inherent limitations of EMR data, which are often unstructured, are likely to contain important omissions

and lack standardized formats. These limitations pose substantial risks to the quality and reproducibility of biomarkers trained on such data. Substantial efforts have been undertaken, in part, by industry and academic collaborations, such as AACR GENIE<sup>198</sup>, and national initiatives, such as Marathon of Hope Cancer Centres Network in Canada<sup>185</sup> and the UK 100,000 Genomes Project<sup>199</sup>, in an attempt to overcome these limitations. To further scale these efforts, collaborative federated learning provides both patient-level security and enables training of computational models across multiple institutions, thus addressing privacy concerns while leveraging large, diverse datasets for model training<sup>184</sup>. However, validation in randomized clinical trials will be crucial for these models to be accepted and implemented in clinical care.

Several technical hurdles limit the reproducibility and robustness of artificial intelligence-based predictive biomarkers, thus complicating their clinical adoption. Efforts by CONSORT-AI and SPIRIT-AI have provided a framework to develop guidance around standards and transparency for clinical trials testing interventions involving artificial intelligence<sup>200</sup>. A lack of initial reproducibility of biomarker associations can arise from the inherent complexity of underlying biological signals, irreproducible research practices or a lack of transparency in methodologies<sup>181</sup>. Guidelines such as CONSORT-AI and SPIRIT-AI provide guidance on appropriate trial designs and reporting procedures to mitigate these challenges<sup>201,202</sup>. Outside of the context of clinical trials, the more recent TRIPOD+AI statement aims to improve the reporting of clinical prediction models to overcome reproducibility-related pitfalls, including enhancing transparency through detailed documentation of data preprocessing, feature engineering, model architecture, hyperparameter tuning and evaluation metrics<sup>203,204</sup>. In addition to these standards, ensuring sufficiently large cohort sizes for both discovery and validation datasets is also important (Table 1) as is appropriate benchmarking to provide accurate and fair assessments of performance, generalizability and clinical utility. Nonetheless, one of the major sources of underperformance associated with predictive models based on curated clinical data is overfitting, which results in deteriorating performance when the model is applied to data beyond the original training dataset<sup>205</sup>. The majority of computationally derived predictive models thus far are based on *in silico* assessments of carefully curated datasets. However, both preclinical and retrospective datasets tend to be biased relative to the intended population, raising concerns of overfitting in predictive models trained on curated data. Preclinical models might also use inappropriate assays or metrics that fail to measure the intended outcomes<sup>206</sup> or outcomes that differ from those investigated in patients (such as the viability of cell lines versus drug responses in patients). Metrics used to assess model performance, such as AUC, often differ from those used to assess biomarker performance in prospective studies, creating a disconnect between research investigations and metrics needed for approval and clinical translation. As a result, biomarkers might seem to perform well in preclinical or retrospective analyses but can have diminished performance when subjected to prospective validation in clinical trials or applied to real-world settings. Any predictive models developed via retrospective analyses need to be rigorously tested in prospective trials, although the lack of representative and high-quality datasets for discovery and retrospective validation increases the risk of unsuccessful prospective validation. Therefore, an imperative need exists – given the potential for life-or-death decisions – for appropriate datasets, experimental and algorithmic designs, and validation cohorts to ensure that predictions are accurate, unbiased and reproducible. Ideally,

these computational biomarkers should be developed alongside the therapeutic product, similarly to companion diagnostics approved alongside specific drugs by major regulatory agencies<sup>207,208</sup>.

Regulations have not kept pace with the development and clinical translation of computational biomarkers, and this mismatch poses one of the most substantial barriers to adoption. Artificial intelligence-based biomarkers, in particular, raise additional regulatory concerns related to algorithmic transparency, interpretability and dataset representativeness. To address these challenges, regulatory standards must evolve beyond conventional performance metrics, such as AUC, sensitivity and specificity, to also require reproducibility and transparency. Given the variability in regulatory approaches across jurisdictions, coordinated international efforts involving regulatory agencies, academic institutions, pharmaceutical and biotech companies, and patient advocacy groups are essential. Organizations such as Friends of Cancer Research are well positioned to spearhead initiatives that advance regulatory policies and define clear, flexible pathways for the evaluation and approval of computational biomarkers. However, even if computational biomarkers demonstrate robust technical performance, given the high resource requirements for implementation and routine use, cost-effectiveness remains an important consideration. Economic feasibility studies for clinical implementation of computational biomarkers are currently lacking; furthermore, little incentive exists for pharmaceutical companies to invest in biomarker development prior to the corresponding treatment being approved. Therefore, to justify integration into clinical workflows, health economics evaluations will be necessary to demonstrate how adoption of computational biomarkers can improve access to precision oncology, reduce costs and ultimately provide benefit to the greatest number of patients.

As the complexity and scale of computational models grow, the adoption and use of these models to match therapies to specific biomarkers remains a critical bottleneck<sup>209</sup>. Less than 40% of oncologists have sufficient training to feel confident in using next-generation sequencing data in their routine practice, let alone being familiar with artificial intelligence and machine learning models<sup>210</sup>. Addressing this gap will require the formation of a genomics-informed and omics-informed healthcare workforce through targeted education and training programmes. Tools such as the ESMO Scale for Clinical Actionability of Molecular Targets<sup>211</sup> and OncoKB<sup>212</sup> provide frameworks to standardize the interpretation of genomic alterations; however, easy-to-use and transparent decision-support tools relating to the use of computational biomarkers will need to be developed and maintained.

## Novel clinical trial designs

Given the potential pitfalls of computational biomarkers described above, prospective trials evaluating these biomarkers will be paramount. Biomarkers derived from *in silico* and retrospective studies, whether computational or molecular, should be considered hypothesis-generating and will require rigorous and continuous clinical validation. Proceeding to clinical implementation without doing so poses a substantial risk of harm to the patient. Improperly validated biomarkers create the risk of denying treatment to patients who might otherwise have benefited (for example, where thresholds for biomarker positivity are set too stringently) or of potential overtreatment and unnecessary toxicities for patients who do not derive benefit (for example, where thresholds for positivity are set overly low). Even more concerning, models trained on misrepresentative data can introduce bias, which

might lead to incorrect treatment recommendations. Furthermore, the process of collecting all the multimodal features can introduce substantial costs and delays in treatment initiation, with the potential for additional harm.

The design of clinical trials incorporating biomarkers poses both logistical and ethical challenges, particularly when stratifying patients on the basis of their genomic and/or functional profiles. The use of patient avatars<sup>56</sup> and digital twins<sup>197,213</sup> can provide an individualized approach for clinical trial matching based on predictions of response and/or toxicities. Innovative clinical trial designs will need to be developed that can enable effective and timely incorporation of computational models as part of biomarker selection. Adaptive trial designs incorporating such models offer potential solutions, for example, by assigning a larger proportion of the participants to biomarker-selected groups that are performing well<sup>214,215</sup>. Notable examples of this approach include the I-SPY 2 trial, which integrates biomarker response data for adaptive randomization to enable individualized assignment of patients to treatment arms, thus maximizing clinical benefit<sup>216</sup>. A real-time adaptive randomization model incorporating continual learning throughout the duration of the trial has also been proposed<sup>217</sup>. Despite the challenges associated with conducting such trials, these novel designs will be necessary to truly determine the extent of benefit from large-scale computational models and to overcome the limitations of conventional trials.

Overcoming these barriers requires multipronged and concerted efforts to standardize methodologies for omics and functional assays, expand open-access databases to democratize data availability, and implement policies that minimize disparities in access and enrolment. We envision multi-data-driven biomarker-informed clinical trials that leverage both clinical and preclinical data to match patients with therapies from which they are most likely to benefit.

## Conclusions

Precision oncology is on the cusp of a new era, with rapid advances in artificial intelligence, computational biology and multi-omics technologies offering unparalleled opportunities to personalize cancer treatment. Yet, translating these innovations into clinical practice remains a formidable challenge, hindered by gaps in data availability, limited model interpretability and an unclear path towards real-world implementation.

An important barrier to progress is the lack of diverse, high-quality datasets that accurately reflect patients with cancer both in clinical trials and real-world settings (Box 1). This issue underscores the urgent need for open science initiatives to democratize access to data and resources. The adoption of the FAIR and IDEA principles is expected not only to accelerate innovation but also to mitigate biases that arise from under-representation of certain populations in model training<sup>182</sup>. Additionally, longitudinal, single-cell and spatially resolved data, such as those derived from liquid biopsies or spatial transcriptomics, are poised to enhance the granularity of predictive biomarkers, capturing features such as tumour dynamics and microenvironmental heterogeneity that bulk and static models fail to address. Given that development of predictive models is increasingly driven by industry, a concerted effort needs to be made to establish shared standards and frameworks – such as the proposed hallmarks for predictive oncology – to guide the development and deployment of these tools<sup>125</sup>. These hallmarks emphasize data and model accessibility, fairness, and clinical applicability, which are essential to fostering trust and enabling widespread adoption.

Despite considerable scientific progress, predictive models remain largely underutilized in clinical practice. This disconnect arises from several factors, including the lack of prospective trials validating their utility, the high costs of implementation, and disparities between the priorities of academic researchers and the commercial imperatives of the pharmaceutical industry. Moreover, clinicians often face challenges in interpreting model outputs owing to limited computational literacy and/or an absence of intuitive decision-support tools. Despite their potential to improve treatment outcomes, resistance to the adoption of artificial intelligence-based models exists owing to various aspects, including the frequent lack of interpretability and risks of biases, which create both ethical and legal challenges. Bridging these gaps will require updated regulatory policies, tailored educational programmes and the development of accessible interfaces that translate complex predictions into explainable and actionable insights for clinicians. Ultimately, overcoming these barriers demands a paradigm shift that aligns the priorities of regulatory agencies, academia, industry and clinical practice. By prioritizing transparency of model development and being inclusive of diverse training data, the field can fulfil its promise to transform cancer care, providing more personalized and effective treatment strategies that will ultimately improve patient outcomes.

Published online: 06 January 2026

## References

- Ashley, E. A. Towards precision medicine. *Nat. Rev. Genet.* **17**, 507–522 (2016).
- Ballman, K. V. Biomarker: predictive or prognostic? *J. Clin. Oncol.* **33**, 3968–3971 (2015).
- Drilon, A. et al. Efficacy of larotrectinib in *TRK* fusion-positive cancers in adults and children. *N. Engl. J. Med.* **378**, 731–739 (2018).
- Meric-Bernstam, F. et al. Efficacy and safety of trastuzumab deruxtecan in patients with HER2-expressing solid tumors: primary results from the DESTINY-PanTumor02 phase II trial. *J. Clin. Oncol.* **42**, 47–58 (2024).
- Hilas, O. Trends in FDA-approved cancer therapies. *U.S. Pharmacist* <https://www.uspharmacist.com/article/trends-in-fdaapproved-cancer-therapies> (2023).
- Mullard, A. NCI-MATCH trial pushes cancer umbrella trial paradigm. *Nat. Rev. Drug Discov.* **14**, 513–515 (2015).
- Tsimberidou, A.-M. et al. Personalized medicine for patients with advanced cancer in the phase I program at MD Anderson: validation and landmark analyses. *Clin. Cancer Res.* **20**, 4827–4836 (2014).
- Stockley, T. L. et al. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med.* **8**, 109 (2016).
- Weiss, F., Lauffenburger, D. & Friedl, P. Towards targeting of shared mechanisms of cancer metastasis and therapy resistance. *Nat. Rev. Cancer* **22**, 157–173 (2022).
- Schmitt, M. W., Loeb, L. A. & Salk, J. J. The influence of subclonal resistance mutations on targeted cancer therapy. *Nat. Rev. Clin. Oncol.* **13**, 335–347 (2016).
- 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).
- Adam, G. et al. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis. Oncol.* **4**, 19 (2020).
- Bhinder, B., Gilvary, C., Madhukar, N. S. & Elemento, O. Artificial intelligence in cancer research and precision medicine. *Cancer Discov.* **11**, 900–915 (2021).
- Tran, K. A. et al. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* **13**, 152 (2021).
- Chiu, Y.-C. et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genomics* **12**, 18 (2019).
- Kuenzi, B. M. et al. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* **38**, 672–684.e6 (2020).
- Alix-Panabières, C. & Pantel, K. Advances in liquid biopsy: from exploration to practical application. *Cancer Cell* **43**, 161–165 (2024).
- Passaro, A. et al. Cancer biomarkers: emerging trends and clinical implications for personalized treatment. *Cell* **187**, 1617–1635 (2024).
- Mateo, J. et al. Delivering precision oncology to patients with cancer. *Nat. Med.* **28**, 658–665 (2022).
- Cai, Z., Poulos, R. C., Liu, J. & Zhong, Q. Machine learning for multi-omics data integration in cancer. *iScience* **25**, 103798 (2022).
- Jee, J. et al. Automated real-world data integration improves cancer outcome prediction. *Nature* **636**, 728–736 (2024).
- Dumbrava, E. E. I. et al. Targeting *ERBB2* (*HER2*) amplification identified by next-generation sequencing in patients with advanced or metastatic solid tumors beyond conventional indications. *JCO Precis. Oncol.* **3**, PO18.00345 (2019).
- Suehnholz, S. P. et al. Quantifying the expanding landscape of clinical actionability for patients with cancer. *Cancer Discov.* **14**, 49–65 (2024).
- Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023).
- Shiao, S. L. et al. Single-cell and spatial profiling identify three response trajectories to pembrolizumab and radiation therapy in triple negative breast cancer. *Cancer Cell* **42**, 70–84.e8 (2024).
- Zuo, C. et al. stClinic dissects clinically relevant niches by integrating spatial multi-slice multi-omics data in dynamic graphs. *Nat. Commun.* **16**, 5317 (2025).
- Cescon, D. W., Bratman, S. V., Chan, S. M. & Siu, L. L. Circulating tumor DNA and liquid biopsy in oncology. *Nat. Cancer* **1**, 276–290 (2020).
- Feng, L. et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit. Health* **4**, e8–e17 (2022).
- Vanguri, R. S. et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat. Cancer* **3**, 1151–1164 (2022).
- Hwang, C.-I., Boj, S. F., Clevers, H. & Tuveson, D. A. Preclinical models of pancreatic ductal adenocarcinoma. *J. Pathol.* **238**, 197–204 (2016).
- Gengenbacher, N., Singhal, M. & Augustin, H. G. Preclinical mouse solid tumour models: status quo, challenges and perspectives. *Nat. Rev. Cancer* **17**, 751–765 (2017).
- Mirabelli, P., Coppola, L. & Salvatore, M. Cancer cell lines are useful model systems for medical research. *Cancers* **11**, 1098 (2019).
- Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4**, 2126 (2013).
- Jin, H. et al. Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nat. Commun.* **14**, 5417 (2023).
- Mouradov, D. et al. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res.* **74**, 3238–3247 (2014).
- Naigebauer, H. et al. CELLector: genomics-guided selection of cancer in vitro models. *Cell Syst.* **10**, 424–432.e6 (2020).
- Warren, C. et al. Global computational alignment of tumor and cell line transcriptional profiles. *Nat. Commun.* **12**, 22 (2021).
- Mourragui, S., Loog, M., van de Wiel, M. A., Reinders, M. J. T. & Wessels, L. F. A. PRECISE: a domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics* **35**, i510–i519 (2019).
- Ma, J. et al. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nat. Cancer* **2**, 233–244 (2021).
- Ashok, G. & Ramaiah, S. A critical review of datasets and computational suites for improving cancer theranostics and biomarker discovery. *Med. Oncol.* **39**, 206 (2022).
- Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
- Iorio, F. et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).
- Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6**, 813–823 (2006).
- Klijn, C. et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).
- Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
- Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Funk, J. S. et al. Deep CRISPR mutagenesis characterizes the functional diversity of TP53 mutations. *Nat. Genet.* **57**, 140–153 (2025).
- Bottomly, D. et al. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell* **40**, 850–864.e9 (2022).
- Pemovska, T. et al. Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. *Nature* **519**, 102–105 (2015).
- Lee, S. et al. High-throughput identification of repurposable neuroactive drugs with potent anti-glioblastoma activity. *Nat. Med.* **30**, 3196–3208 (2024).
- Acanda De La Rocha, A. M. et al. Feasibility of functional precision medicine for guiding treatment of relapsed or refractory pediatric cancers. *Nat. Med.* **30**, 990–1000 (2024).
- Dolgin, E. The future of precision cancer therapy might be to try everything. *Nature* **626**, 470–473 (2024).
- Kytölä, S. et al. Ex vivo venetoclax sensitivity predicts clinical response in acute myeloid leukemia in the prospective VenEx trial. *Blood* **145**, 409–421 (2025).
- Lee, S. H. R. et al. Pharmacotypes across the genomic landscape of pediatric acute lymphoblastic leukemia and impact on treatment response. *Nat. Med.* **29**, 170–179 (2023).
- Sarno, F. et al. A phase III randomized trial of integrated genomics and avatar models for personalized treatment of pancreatic cancer: the AVATAR trial. *Clin. Cancer Res.* **31**, 278–287 (2025).
- Ayala Solares, J. R. et al. Deep learning for electronic health records: a comparative review of multiple deep neural architectures. *J. Biomed. Inf.* **101**, 103337 (2020).

58. Barche, Y. et al. Leveraging big data of immune checkpoint blockade response identifies novel potential targets. *Ann. Oncol.* **33**, 1304–1317 (2022).
59. Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614.e14 (2021).
60. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
61. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
62. Hoppe, M. M., Sundar, R., Tan, D. S. P. & Jeyasekharan, A. D. Biomarkers for homologous recombination deficiency in cancer. *J. Natl Cancer Inst.* **110**, 704–713 (2018).
63. Telli, M. L. et al. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin. Cancer Res.* **22**, 3764–3773 (2016).
64. Figueroa, M. E. et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).
65. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
66. Macintyre, G. et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1262–1270 (2018).
67. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
68. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
69. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
70. Billan, S., Kaidar-Person, O. & Gil, Z. Treatment after progression in the era of immunotherapy. *Lancet Oncol.* **21**, e463–e476 (2020).
71. Borcoman, E. et al. Novel patterns of response under immunotherapy. *Ann. Oncol.* **30**, 385–396 (2019).
72. Nathan, P. et al. Overall survival benefit with tivantinib in metastatic uveal melanoma. *N. Engl. J. Med.* **385**, 1196–1206 (2021).
73. van 't Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
74. Cardoso, F. et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.* **375**, 717–729 (2016).
75. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
76. Zhao, E. Y. et al. Homologous recombination deficiency and platinum-based therapy outcomes in advanced breast cancer. *Clin. Cancer Res.* **23**, 7521–7530 (2017).
77. Patsouris, A. et al. Rucaparib in patients presenting a metastatic breast cancer with homologous recombination deficiency, without germline BRCA1/2 mutation. *Eur. J. Cancer* **159**, 283–295 (2021).
78. Smirnov, P. et al. Evaluation of statistical approaches for association testing in noisy drug screening data. *BMC Bioinformatics* **23**, 188 (2022).
79. Anagnostou, V. et al. Multimodal genomic features predict outcome of immune checkpoint blockade in non-small-cell lung cancer. *Nat. Cancer* **1**, 99–111 (2020).
80. Slamon, D. J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182 (1987).
81. Maemondo, M. et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N. Engl. J. Med.* **362**, 2380–2388 (2010).
82. Flaherty, K. T. et al. Improved survival with MEK inhibition in BRAF-mutated melanoma. *N. Engl. J. Med.* **367**, 107–114 (2012).
83. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
84. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
85. Filipits, M. et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin. Cancer Res.* **17**, 6012–6020 (2011).
86. Christopoulos, P. et al. Plasma proteome-based test for first-line treatment selection in metastatic non-small cell lung cancer. *JCO Precis. Oncol.* **8**, e2300555 (2024).
87. Ali, M. & Aittokallio, T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys. Rev.* **11**, 31–39 (2019).
88. Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* **23**, 40–55 (2022).
89. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592 (2018).
90. Sharifi-Noghabi, H. et al. Drug sensitivity prediction from cell line-based pharmacogenomics data: guidelines for developing machine learning models. *Brief. Bioinform.* **22**, bbab294 (2021).
91. Ishwaran, H., Kogalur, U. B., Blackstone, E. H. & Lauer, M. S. Random survival forests. *Ann. Appl. Stat.* **2**, 841–860 (2008).
92. Pölslerl, S., Navab, N. & Katouzian, A. In: *Machine Learning and Knowledge Discovery in Databases* (eds Appice, A. et al.) 243–259 (Springer, 2015).
93. Unger, M. & Kather, J. N. A systematic analysis of deep learning in genomics and histopathology for precision oncology. *BMC Med.* **17**, 48 (2024).
94. Katzman, J. L. et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **18**, 24 (2018).
95. Schmauch, B. et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat. Commun.* **11**, 3877 (2020).
96. Hoang, D.-T. et al. Prediction of DNA methylation-based tumor types from histopathology in central nervous system tumors with deep learning. *Nat. Med.* **30**, 1952–1961 (2024).
97. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
98. Fu, Y. et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* **1**, 800–810 (2020).
99. Kather, J. N. et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* **1**, 789–799 (2020).
100. Hoang, D.-T. et al. A deep-learning framework to predict cancer treatment response from histopathology images through imputed transcriptomics. *Nat. Cancer* **5**, 1305–1317 (2024).
101. Vaswani, A. et al. Attention is all you need. In *Proc. 31st Conference on Neural Information Processing Systems 6000–6010* (NeurIPS, 2017).
102. Yang, X. et al. A large language model for electronic health records. *NPJ Digit. Med.* **5**, 194 (2022).
103. Pai, S. et al. Foundation model for cancer imaging biomarkers. *Nat. Mach. Intell.* **6**, 354–367 (2024).
104. Chen, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nat. Med.* **30**, 850–862 (2024).
105. Xu, H. et al. A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**, 181–188 (2024).
106. Wang, X. et al. A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* **634**, 970–978 (2024).
107. Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 (2024).
108. Theodoris, C. V. et al. Transfer learning enables predictions in network biology. *Nature* **618**, 616–624 (2023).
109. Hao, M. et al. Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* **21**, 1481–1491 (2024).
110. Wang, C. et al. ScGPT-spatial: continual pretraining of single-cell foundation model for spatial transcriptomics. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.02.05.636714> (2025).
111. Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
112. van de Haar, J. et al. Combining genomic biomarkers to guide immunotherapy in non-small cell lung cancer. *Clin. Cancer Res.* **30**, 1307–1318 (2024).
113. Usset, J. et al. Five latent factors underlie response to immunotherapy. *Nat. Genet.* **56**, 2112–2120 (2024).
114. Chang, T.-G. et al. LORIS robustly predicts patient outcomes with immune checkpoint blockade therapy using common clinical, pathologic and genomic features. *Nat. Cancer* **5**, 1158–1175 (2024).
115. Steyaert, S. et al. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat. Mach. Intell.* **5**, 351–362 (2023).
116. Lipkova, J. et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**, 1095–1110 (2022).
117. Carrillo-Perez, F. et al. Machine-learning-based late fusion on multi-omics and multi-scale data for non-small-cell lung cancer diagnosis. *J. Pers. Med.* **12**, 601 (2022).
118. Cheerla, A. & Gevaert, O. Deep learning with multimodal representation for pancreatic prognosis prediction. *Bioinformatics* **35**, i446–i454 (2019).
119. Ma, S. et al. Moving towards genome-wide data integration for patient stratification with Integrate Any Omics. *Nat. Mach. Intell.* **7**, 29–42 (2025).
120. Xu, P., Zhu, X. & Clifton, D. A. Multimodal learning with Transformers: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 12113–12132 (2023).
121. Xiang, J. et al. A vision-language foundation model for precision oncology. *Nature* **638**, 769–778 (2025).
122. Cui, H. et al. Towards multimodal foundation models in molecular cell biology. *Nature* **640**, 623–633 (2025).
123. Arango-Argoty, G. et al. Pretrained transformers applied to clinical studies improve predictions of treatment efficacy and associated biomarkers. *Nat. Commun.* **16**, 2101 (2025).
124. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126 (2022).
125. Singhal, A. et al. The hallmarks of predictive oncology. *Cancer Discov.* **15**, 271–285 (2025).
126. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W. & O'Sullivan, J. M. A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinform.* **2**, 927312 (2022).
127. Cao, Y., Geddes, T. A., Yang, J. Y. H. & Yang, P. Ensemble deep learning in bioinformatics. *Nat. Mach. Intell.* **2**, 500–508 (2020).
128. Sammut, S.-J. et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* **601**, 623–629 (2022).
129. Zhang, Y. & Yang, Q. An overview of multi-task learning. *Natl Sci. Rev.* **5**, 30–43 (2018).
130. Thung, K.-H. & Wee, C.-Y. A brief review on multi-task learning. *Multimed. Tools Appl.* **77**, 29705–29725 (2018).
131. White, B. S. et al. Bayesian multi-source regression and monocyte-associated gene expression predict BCL-2 inhibitor resistance in acute myeloid leukemia. *NPJ Precis. Oncol.* **5**, 71 (2021).
132. Costello, J. C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).
133. Kowald, A. et al. Transfer learning of clinical outcomes from preclinical molecular data, principles and perspectives. *Brief. Bioinform.* **23**, bbab133 (2022).

134. Kirkpatrick, J. et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl Acad. Sci. USA* **114**, 3521–3526 (2017).
135. Li, Z. & Hoiem, D. Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 2935–2947 (2018).
136. Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T. P. & Wayne, G. Experience replay for continual learning. In *Proc. 33rd Conference on Neural Information Processing Systems* 350–360 (NeurIPS 2019).
137. Rusu, A. A. et al. Progressive neural networks. Preprint at <https://doi.org/10.48550/arXiv.1606.04671> (2016).
138. Lopez-Paz, D. & Ranzato, M. Gradient episodic memory for continual learning. In *Proc. 31st International Conference on Neural Information Processing Systems* 6470–6479 (NeurIPS, 2017).
139. Xu, Y. et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin. Cancer Res.* **25**, 3266–3275 (2019).
140. Kummur, S. et al. Utilizing targeted cancer therapeutic agents in combination: novel approaches and urgent requirements. *Nat. Rev. Drug Discov.* **9**, 843–856 (2010).
141. Jin, H., Wang, L. & Bernards, R. Rational combinations of targeted cancer therapies: background, advances and challenges. *Nat. Rev. Drug Discov.* **22**, 213–234 (2023).
142. Conroy, T. et al. FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *N. Engl. J. Med.* **379**, 2395–2406 (2018).
143. Von Hoff, D. D. et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N. Engl. J. Med.* **369**, 1691–1703 (2013).
144. Schmid, P. et al. Pembrolizumab for early triple-negative breast cancer. *N. Engl. J. Med.* **382**, 810–821 (2020).
145. Antonia, S. J. et al. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. *N. Engl. J. Med.* **379**, 2342–2350 (2018).
146. Loupakis, F. et al. Initial therapy with FOLFIRI and bevacizumab for metastatic colorectal cancer. *N. Engl. J. Med.* **371**, 1609–1618 (2014).
147. Jaaks, P. et al. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature* **603**, 166–173 (2022).
148. Sicklick, J. K. et al. Molecular profiling of advanced malignancies guides first-line N-of-1 treatments in the I-PREDICT treatment-naïve study. *Genome Med.* **13**, 155 (2021).
149. Sicklick, J. K. et al. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat. Med.* **25**, 744–750 (2019).
150. Meric-Bernstam, F. et al. National Cancer Institute combination therapy platform trial with molecular analysis for therapy choice (ComboMATCH). *Clin. Cancer Res.* **29**, 1412–1422 (2023).
151. Li, X. et al. Precision combination therapies based on recurrent oncogenic coalterations. *Cancer Discov.* **12**, 1542–1559 (2022).
152. US National Library of Medicine. *ClinicalTrials.gov* <https://clinicaltrials.gov/study/NCT05674825> (2025).
153. Menden, M. P. et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat. Commun.* **10**, 2674 (2019).
154. Kong, W. et al. Systematic review of computational methods for drug combination prediction. *Comput. Struct. Biotechnol. J.* **20**, 2807–2814 (2022).
155. Dinstag, G. et al. Clinically oriented prediction of patient response to targeted and immunotherapies from the tumor transcriptome. *Med* **4**, 15–30.e8 (2023).
156. Kuru, H. I., Cicek, A. E. & Tastan, O. From cell lines to cancer patients: personalized drug synergy prediction. *Bioinformatics* **40**, btac134 (2022).
157. Chen, Y. et al. A machine learning-based strategy predicts selective and synergistic drug combinations for relapsed acute myeloid leukemia. *Cancer Res.* **85**, 2753–2768 (2025).
158. Pomeroy, A. E., Schmidt, E. V., Sorger, P. K. & Palmer, A. C. Drug independence and the curability of cancer by combination chemotherapy. *Trends Cancer* **8**, 915–929 (2022).
159. Bunne, C. et al. Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* **20**, 1759–1768 (2023).
160. Roohani, Y., Huang, K. & Leskovec, J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nat. Biotechnol.* **42**, 927–935 (2024).
161. Zhang, W. et al. Integration of pan-cancer cell line and single-cell transcriptomic profiles enables inference of therapeutic vulnerabilities in heterogeneous tumors. *Cancer Res.* **84**, 2021–2033 (2024).
162. Schrod, S., Zacharias, H. U., Beifbarth, T., Hauschild, A.-C. & Altenbuchinger, M. CODEX: counterfactual deep learning for the in silico exploration of cancer cell line perturbations. *Bioinformatics* **40**, i91–i99 (2024).
163. Maeser, D., Zhang, W., Huang, Y. & Huang, R. S. A review of computational methods for predicting cancer drug response at the single-cell level through integration with bulk RNAseq data. *Curr. Opin. Struct. Biol.* **84**, 102745 (2024).
164. Ianevski, A. et al. Patient-tailored design for selective co-inhibition of leukemic cell subpopulations. *Sci. Adv.* **7**, eabe4038 (2021).
165. Sinha, S. et al. PERCEPTION predicts patient response and resistance to treatment using single-cell transcriptomics of their tumors. *Nat. Cancer* **5**, 938–952 (2024).
166. Ianevski, A. et al. Single-cell transcriptomes identify patient-tailored therapies for selective co-inhibition of cancer clones. *Nat. Commun.* **15**, 8579 (2024).
167. Kornauth, C. et al. Functional precision medicine provides clinical benefit in advanced aggressive hematologic cancers and identifies exceptional responders. *Cancer Discov.* **12**, 372–387 (2022).
168. Miglino, N. et al. Feasibility of multiomics tumor profiling for guiding treatment of melanoma. *Nat. Med.* **31**, 2430–2441 (2025).
169. Sheridan, C. Can single-cell biology realize the promise of precision medicine? *Nat. Biotechnol.* **42**, 159–162 (2024).
170. PromethION. *Oxford Nanopore Technologies* <https://nanoporetech.com/products/sequence/promethion> (2025).
171. Tang, C. et al. Personalized tumor combination therapy optimization using the single-cell transcriptome. *Genome Med.* **15**, 105 (2023).
172. Tatarova, Z. et al. A multiplex implantable microdevice assay identifies synergistic combinations of cancer immunotherapies and conventional drugs. *Nat. Biotechnol.* **40**, 1823–1833 (2022).
173. Lundberg, S. & Lee, S.-I. A unified approach to interpreting model predictions. In *Proc. 31st International Conference on Neural Information Processing Systems* 4765–4774 (2017).
174. Ribeiro, M. T., Singh, S. & Guestrin, C. Why should I trust you?: explaining the predictions of any classifier. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Vol. 16, 1135–1144 (Association for Computing Machinery, 2016).
175. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2921–2929 (IEEE, 2016).
176. Tomaszewski, M. R. & Gillies, R. J. The biological meaning of radiomic features. *Radiology* **298**, 505–516 (2021).
177. Ancona, M., Ceolini, E., Öztireli, C. & Gross, M. Towards better understanding of gradient-based attribution methods for deep neural networks. In *Proc. 6th International Conference on Learning Representations* <https://doi.org/10.3929/ethz-b-000249929> (ICLR, 2018).
178. Shrikumar, A., Greenside, P. & Kundaje, A. Learning important features through propagating activation differences. In *Proc. 34th International Conference on Machine Learning* 3145–3153 (PMLR, 2017).
179. Elmarakeby, H. A. et al. Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348–352 (2021).
180. Park, S. et al. A deep learning model of tumor cell architecture elucidates response and resistance to CDK4/6 inhibitors. *Nat. Cancer* **5**, 996–1009 (2024).
181. Haibe-Kains, B. et al. Transparency and reproducibility in artificial intelligence. *Nature* **586**, E14–E16 (2020).
182. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
183. Rieke, N. et al. The future of digital health with federated learning. *NPJ Digit. Med.* **3**, 119 (2020).
184. Ogier du Terrail, J. et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat. Med.* **29**, 135–146 (2023).
185. Terry Fox Research Institute Marathon of Hope Cancer Centers Network. The Terry Fox Research Institute Marathon of Hope Cancer Centres Network: a pan-Canadian precision oncology initiative. *Cancer Cell* **43**, 587–592 (2025).
186. Mahon, P. et al. A federated learning system for precision oncology in Europe: DigiONE. *Nat. Med.* **30**, 334–337 (2024).
187. Wen, J. et al. A survey on federated learning: challenges and applications. *Int. J. Mach. Learn. Cybern.* **14**, 513–535 (2023).
188. Moon, I. et al. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. *Nat. Med.* **29**, 2057–2067 (2023).
189. Ianni, J. D. et al. Tailored for real-world: a whole slide image classification system validated on uncured multi-site data emulating the prospective pathology workload. *Sci. Rep.* **10**, 3217 (2020).
190. Feuerriegel, S. et al. Causal machine learning for predicting treatment outcomes. *Nat. Med.* **30**, 958–968 (2024).
191. Lone, S. N. et al. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Mol. Cancer* **21**, 79 (2022).
192. Sanz-Garcia, E., Zhao, E., Bratman, S. V. & Siu, L. L. Monitoring and adapting cancer treatment using circulating tumor DNA kinetics: current research, opportunities, and challenges. *Sci. Adv.* **8**, eabi8618 (2022).
193. Rozenblatt-Rosen, O. et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249 (2020).
194. Frampton, G. M. et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023–1031 (2013).
195. Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L. & Siu, L. L. Molecular profiling for precision cancer therapies. *Genome Med.* **12**, 8 (2020).
196. Cai, Z. et al. Synthetic augmentation of cancer cell line multi-omic datasets using unsupervised deep learning. *Nat. Commun.* **15**, 10390 (2024).
197. Thangaraj, P. M., Shankar, S. V., Oikonomou, E. K. & Khera, R. RCT-Twin-GAN generates digital twins of randomized control trials adapted to real-world patients to enhance their inference and application. Preprint at [medRxiv https://doi.org/10.1101/2023.12.06.23299464](https://doi.org/10.1101/2023.12.06.23299464) (2023).
198. AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discov.* **7**, 818–831 (2017).
199. Turnbull, C. et al. The 100 000 genomes project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).
200. Topol, E. J. Welcoming new guidelines for AI clinical research. *Nat. Med.* **26**, 1318–1320 (2020).
201. Liu, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat. Med.* **26**, 1364–1374 (2020).
202. Cruz Rivera, S. et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat. Med.* **26**, 1351–1363 (2020).
203. Collins, G. S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **385**, e078378 (2024).
204. Gallifant, J. et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat. Med.* **31**, 60–69 (2025).

205. Garassino, M. C. et al. PLO2.11 Normalized membrane ratio of TROP2 by quantitative continuous scoring is predictive of clinical outcomes in TROPION-Lung 01. *J. Thorac. Oncol.* **19**, S2–S3 (2024).
206. Foy, R., Lew, K. X. & Saurin, A. T. The search for CDK4/6 inhibitor biomarkers has been hampered by inappropriate proliferation assays. *NPJ Breast Cancer* **10**, 19 (2024).
207. U.S. Food and Drug Administration. *Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product* <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-codevelopment-in-vitro-companion-diagnostic-device-therapeutic-product> (2019).
208. Maliepaard, M., Nibi, P., Nibi, G. & Pasmooij, A. M. G. Evaluation of companion diagnostics in scientific advice and drug marketing authorization applications by the European Medicines Agency. *Front. Med.* **9**, 893028 (2022).
209. Horgan, D. et al. Bringing greater accuracy to Europe's healthcare systems: the unexploited potential of biomarker testing in oncology. *Biomed. Hub.* **5**, 182–223 (2020).
210. de Moor, J. S., Gray, S. W., Mitchell, S. A., Klabunde, C. N. & Freedman, A. N. Oncologist confidence in genomic testing and implications for using multimer tumor panel tests in practice. *JCO Precis. Oncol.* **4**, 620–631 (2020).
211. Mateo, J. et al. A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). *Ann. Oncol.* **29**, 1895–1902 (2018).
212. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00011> (2017).
213. Asghar, U. S. & Chung, C. Application of digital twins for personalized oncology. *Nat. Rev. Cancer* **25**, 823–825 (2025).
214. Bhatt, D. L. & Mehta, C. Adaptive designs for clinical trials. *N. Engl. J. Med.* **375**, 65–74 (2016).
215. Thorlund, K., Haggstrom, J., Park, J. J. & Mills, E. J. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ* **360**, k698 (2018).
216. Wang, H. & Yee, D. I-SPY 2: a neoadjuvant adaptive clinical trial designed to improve outcomes in high-risk breast cancer. *Curr. Breast Cancer Rep.* **11**, 303–310 (2019).
217. Marret, G., Herrera, M. & Siu, L. L. Turning the kaleidoscope: innovations shaping the future of clinical trial design. *Cancer Cell* **43**, 597–605 (2025).
218. Drukker, C. A. et al. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int. J. Cancer* **133**, 929–936 (2013).
219. Sparano, J. A. et al. Prospective validation of a 21-gene expression assay in breast cancer. *N. Engl. J. Med.* **373**, 2005–2014 (2015).
220. Alvarez, M. J. et al. A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat. Genet.* **50**, 979–989 (2018).
221. Kather, J. N. et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
222. Mu, W. et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J. Immunother. Cancer* **9**, e002118 (2021).
223. Wagner, S. J. et al. Transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. *Cancer Cell* **41**, 1650–1661.e4 (2023).
224. Sun, R. et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* **19**, 1180–1191 (2018).
225. Preuer, K. et al. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* **34**, 1538–1546 (2018).
226. Xia, F. et al. Predicting tumor cell line response to drug pairs with deep learning. *BMC Bioinformatics* **19**, 486 (2018).
227. Zhang, T., Zhang, L., Payne, P. R. O. & Li, F. Synergistic drug combination prediction by integrating multiomics data in deep learning models. *Methods Mol. Biol.* **2194**, 223–238 (2021).
228. Kim, Y. et al. Anticancer drug synergy prediction in understudied tissues using transfer learning. *J. Am. Med. Inf. Assoc.* **28**, 42–51 (2021).
229. Yang, J., Xu, Z., Wu, W. K. K., Chu, Q. & Zhang, Q. GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J. Am. Med. Inf. Assoc.* **28**, 2336–2345 (2021).
230. Liu, Q. & Xie, L. TranSynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput. Biol.* **17**, e1008653 (2021).
231. Kuru, H. I., Tastan, O. & Cicek, A. E. MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**, 2334–2344 (2022).
232. Wang, X. et al. PRODeepSyn: predicting anticancer synergistic drug combinations by embedding cell lines with protein-protein interaction network. *Brief. Bioinform.* **23**, bbab587 (2022).
233. Hu, J. et al. DTSyn: a dual-transformer-based neural network to predict synergistic drug combinations. *Brief. Bioinform.* **23**, bbac302 (2022).
234. Tang, Y.-C. & Gottlieb, A. SynPathy: predicting drug synergy through drug-associated pathways using deep learning. *Mol. Cancer Res.* **20**, 762–769 (2022).
235. Hosseini, S.-R. & Zhou, X. CCSynergy: an integrative deep-learning framework enabling context-aware prediction of anti-cancer drug synergy. *Brief. Bioinform.* **24**, bbac588 (2023).
236. Xu, M. et al. DFFNDDS: prediction of synergistic drug combinations with dual feature fusion networks. *J. Cheminform.* **15**, 33 (2023).
237. El Khili, M. R., Memon, S. A. & Emad, A. MARSy: a multitask deep-learning framework for prediction of drug combination synergy scores. *Bioinformatics* **39**, btad177 (2023).
238. Zhang, P. & Tu, S. MGAE-DC: predicting the synergistic effects of drug combinations through multi-channel graph autoencoders. *PLoS Comput. Biol.* **19**, e1010951 (2023).
239. Li, T.-H., Wang, C.-C., Zhang, L. & Chen, X. SNRMPACDC: computational model focused on Siamese network and random matrix projection for anticancer synergistic drug combination prediction. *Brief. Bioinform.* **24**, bbac503 (2023).
240. Verkerk, K. & Voest, E. E. Generating and using real-world data: a worthwhile uphill battle. *Cell* **187**, 1636–1650 (2024).
241. Yu, Y. et al. Association of survival and immune-related biomarkers with immunotherapy in patients with non-small cell lung cancer: a meta-analysis and individual patient-level analysis. *JAMA Netw. Open* **2**, e196879 (2019).

## Acknowledgements

We sincerely thank L. Siu of the Princess Margaret Cancer Centre for her insightful comments on this Review. The authors' research is supported by funding from the ARPA-H ADAPT Program (OT 140D042590013 to T.I. and B.H.-K.).

## Author contributions

X.W. and J.N. initiated and coordinated the Review. X.W., J.N., K.N., M.M., P.W., A.S., T.A. and B.H.-K. researched data for the manuscript and made a substantial contribution to discussions of content. X.W. and P.L.B. provided clinical input. T.I., T.A. and B.H.-K. supervised the Review. All authors edited and/or reviewed the manuscript prior to submission.

## Competing interests

P.L.B. reports research funding (to the institution) from Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Bicara Therapeutics, Bristol Myers Squibb, Daiichi Sankyo, Genentech/Roche, GlaxoSmithKline, LegoChem, Lilly, Merck, Medicenna, Novartis, Seagen/Pfizer, Takeda and Zymeworks. He also reports uncompensated honoraria/consultancy with Amgen, Daiichi Sankyo, Gilead, Janssen, Lilly, Merck, Repare, Seattle Genetics and Zymeworks, and a compensated advisory role for Boehringer Ingelheim. T.I. is a co-founder, member of the advisory board, and has an equity interest in Data4Cure and Serinus Biosciences, is a consultant of and has an equity interest in Ideaya Biosciences, and is a member of the advisory editorial board of *The EMBO Journal*. The terms of all consultancy and equity arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies. B.H.-K. is part of the scientific advisory boards of Break Through Cancer, Commonwealth Cancer Consortium (USA), Consortium de Recherche Biopharmaceutique (CQDM; Quebec, Canada), the Canadian Institute of Health Research — Institute of Genetics, Cancer Grand Challenges (UK) and Chriners Children (Florida, USA), is a co-founder of the MAQC (Massive Analysis and Quality Control) Society and is part of the board of directors of AACR International (Canada) and The American Association for Cancer Research (USA). X.W., J.N., K.N., M.M., P.W., A.S. and T.A. declare no competing interests.

## Additional information

**Peer review information** *Nature Reviews Clinical Oncology* thanks C. Cheng, V. Fortino, J. Kather and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2026

<sup>1</sup>Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. <sup>2</sup>Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. <sup>3</sup>Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. <sup>4</sup>iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>5</sup>Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. <sup>6</sup>Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA. <sup>7</sup>Department of Medicine, University of California, San Diego, La Jolla, CA, USA. <sup>8</sup>Moore's Cancer Center, University of California, San Diego, San Diego, CA, USA. <sup>9</sup>Institute for Cancer Research, Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway. <sup>10</sup>Oslo Centre for Biostatistics and Epidemiology (OCBE), Faculty of Medicine, University of Oslo, Oslo, Norway. <sup>11</sup>Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada. <sup>12</sup>Structural Genomics Consortium, University Health Network, Toronto, Ontario, Canada. <sup>13</sup>These authors contributed equally: Xin Wang, Julia Nguyen.