ELSEVIER

KEYNOTE (GREEN)

Causal inference in drug discovery and development

Tom Michoel [1], **Jitao David Zhang** [2,3,*]

[1] Computational Biology Unit, Department of Informatics, University of Bergen, Postboks 7803, 5020 Bergen, Norway
[2] Pharma Early Research and Development, Roche Innovation Centre Basel, F. Hoffmann-La Roche, Grenzacherstrasse 124, 4070 Basel, Switzerland
[3] Department of Mathematics and Computer Science, University of Basel, Spiegelgasse 1, 4051 Basel, Switzerland

Tom Michoel is professor in bioinformatics at the University of Bergen. His research focuses on developing methods for causal inference from multi-omics data.

Jitao David Zhang is a computational biologist working in drug discovery. His research focuses on multiscale-scale causal modelling for risk-benefit assessment of drug candidates.

To discover new drugs is to seek and to prove causality. As an emerging approach leveraging human knowledge and creativity, data, and machine intelligence, causal inference holds the promise of reducing cognitive bias and improving decision-making in drug discovery. Although it has been applied across the value chain, the concepts and practice of causal inference remain obscure to many practitioners. This article offers a nontechnical introduction to causal inference, reviews its recent applications, and discusses opportunities and challenges of adopting the causal language in drug discovery and development.

Keywords: causality; causal inference; DAG (standing for directed acyclic graph); drug discovery; drug development; reverse translation

Introduction

Causal inference is the process of identifying causal effects based on prior knowledge, hypothesis, and correlations observed in data. This article aims at equipping practitioners of drug discovery and development, especially those without formal training in mathematics and statistics, with necessary knowledge to start working with causal inference at three levels: (i) to recognize situations where causal inference is advantageous and to understand why a causal model is needed (Section *What is causal inference and why do we need it?*), (ii) to perform causal inference with software (Section *A step-by-step guide of causal inference*), and (iii) to learn and get inspiration from recent applications (Section *Literature review and case studies*). Finally, we discuss opportunities and challenges of adopting the causal language in drug discovery and development. A glossary is offered in the Supplementary File 1 in the Supplementary material online for readers to look up definitions of important terms.

Causal inference identifies causations from correlations. In statistics, *correlation* means the relationship between two random variables, whether causal or not. For instance, high temperatures lead to higher ice cream sales; therefore, the variables *temperature* and *ice cream sales* are correlated. Similarly, high temperatures make wildfire more likely to happen; therefore, the variables *temperature* and *wildfire frequency* are correlated. Apparently, *ice cream sales* and *wildfire frequency* are correlated but do not cause each other. The task of causal inference in this context is to infer

* *Corresponding author. Zhang, J.D. (jitao_david.zhang@roche.com)*

the causal relationship between *temperature, ice cream sales,* and *wildfire frequency* based on observations.

There is neither a consensus definition of causality nor a single way to identify it (Box 1). It is partially because the nature of causation is still under debate,[1] and partially because even simple causality such as *higher temperature causes higher ice cream sales* can be broken down into an infinite chain of causal relationships that can involve entities across both physical (e.g., molecules, receptors, cells, organs and systems, society) and time scales.

In this article, we introduce the statistical causal inference approach and define *causality* as a probabilistic relationship that satisfies four conditions: *regular probabilistic update*, *manipulation*, *counterfactual condition*, and *mechanism of action*.

Take the example of testing whether a drug is causal for halting deadly cancer progression. *Regular probabilistic update* means that taking the drug modifies the conditional probability of dying of the disease within a defined time window, regardless of where and when the trial happens. *Manipulation* means that drug treatment shows additional benefit even if we consider all other factors affecting patient survival, such as age and comorbidities. *Counterfactual condition* means that the death of a patient would not have been postponed had the drug not been taken. Finally, *mechanism of action* means that we understand why the drug prolongs patient survival, such as by activating tumour-infiltrating immune cells. Taken together, the four conditions ensure both statistical correlation and mechanistic understanding. They embody the causality criteria proposed by Austin Bradford Hill[2] and implement a practical test for the philosophical reasoning of establishing causality in healthcare.[3]

What is causal inference and why do we need it?

Why causality matters in drug discovery and development

Causality is indispensable for predicting outcomes of intervention, for answering counterfactual *what if* questions, and for human understanding. Although it takes time and effort to identify causes from correlations, the investment is rewarding. The history of drug discovery abounds with cases where causality inspires repurposing of a drug, developing new classes of chemical matters, or refuting the use of a drug.

One prominent example is the history of thalidomide.[4] Chemically, thalidomide is a mixture of two enantiomers: the sedative (*R*)-enantiomer and the teratogenic (*S*)-enantiomer. Although its use was first correlated with an unusually high incidence of birth defects,[5] it took decades to identify the cause of the teratogenicity, that is, the degradation of SALL4 protein via the E3 ubiquitin ligase complex.[6] In the meantime, thalidomide was found to be potentially effective against drug-resistant multiple myeloma. Follow-up study based on the first causal finding offered a causal explanation for this correlation, namely that thalidomide induces protein degradation of key transcription factors *Ikaros* (IKZF1) and *Aiolos* (IKZF3) with the same ubiquitin apparatus. Furthermore, the same causal link between thalidomide and protein degradation has inspired a flourishing search for new modalities, including proteolysis-targeting chimera (PROTAC)[7] and other multispecific drugs.[8] The ups and downs of thalidomide, in which causality plays a central role in translating serendipitous findings into new discoveries, is therefore described as Shakespearean.[4]

Repurposing drugs to treat COVID-19 offers another recent example demonstrating the importance of causality. Much literature reported a correlation between treatment with drugs targeting one or both human sigma receptors (σ_1 and σ_2), such as hydroxychloroquine, and negative modulation of SARS-CoV-2 infection. Researchers, however, noticed the intriguing discrepancy that although most compounds have comparable affinity against sigma receptors, the cellular antiviral activities showed large variation. It raised the question whether pharmacological modulation of sigma receptors is causal for SARS-CoV-2 inhibition. The question led Tummino *et al.*[9] to find that phospholipidosis, instead of target-mediated mechanisms, underlies the antiviral activity of many drugs. Phospholipidosis, determined by physicochemical properties of the drugs, results in dose-limiting toxicity and can be reliably predicted experimentally and computationally.[10] The lack of causal relationship between pharmacological modulation of sigma receptors and antiviral activities highlights the importance of dissecting causal mechanisms from correlations, from the opposite angle of the example of thalidomide.

Although correlations suffice for predictions and (wrong) explanations, causality enables us to understand why a drug works or causes harm and to act rationally. What is the relationship between correlation and causation? How can we systematically identify causations from correlations?

Distinguishing causation from correlation

Causality is often hidden and manifests itself by correlations, though correlations are not always due to direct causality. There are six reasons why we observe a correlation between two events: *causation*, *confounding*, *coincidence*, *conspiracy*, *collider,* and *chronology*.

We assume that two variables, x and y, depict expression levels of two proteins, X and Y, in a population of cells. If x and y are correlated, six scenarios are possible:

1. Causation: expression of X causes expression of Y, or expression of Y causes expression of X. We may use additional knowledge to favour one direction over the other. For instance, if X is a known transcription factor but Y is not, we may favour the model X → Y (→reads causes) over Y → X.
2. Confounding: a third, potentially unobserved, protein U causes expression of both X and Y, i.e. X ← U → Y.
3. Coincidence: the correlation is solely by chance. If so, we shall observe diminishing correlations as we collect more data. We can perform statistical inferences, such as permutation test and bootstrapping, to test how likely we observe the correlation by chance.
4. Conspiracy, or deliberate selection of data: the correlation is due to deliberate manipulation of data or the sampling process. We may, for instance, create a good correlation by removing data from all cells where two proteins are not correlated.
5. Collider: besides conspiracy, nondeliberate selection of observations from the general population can also cause correlation. Even if x and y are not correlated in all cells (imagine

random points confined in a circle), by considering only cells in which at least one of two proteins is expressed over a certain threshold (imagine removing points between six and nine o'clock), we observe a negative correlation between x and y. This is known as the Berkson's paradox[11] or the admission rate bias.[12]

6. Chronology: If both X and Y change their expression during the life cycle of a cell, they are correlated in time series even if neither protein causes change of expression of the other. Although time can be considered a confounder, it is not a cause per se of either protein's expression. Such independent trends with regard to time are systematically explored by Hans Reichenbach in The Direction of Time.[13] This is a subject of intensive debate in contemporary philosophy.[14]

Collectively, we call the six mechanisms the *6C model of correlation*. The model helps us distil real causes from empirical correlations by excluding the possibility that alternative mechanisms are at work. In particular, data and statistical models alone cannot tell causation apart from confounding. Figure 1 illustrates a 'toy' example in the same line of the protein expression example above. It demonstrates that in order to identify causal relationships from correlations, we need a model of either knowledge or hypotheses about how data is generated and to operate the model with statistical modelling.

For some applications, correlation is all we need. In the toy example illustrated, if our goal is to predict y given x in similar yet unobserved cells of the same population, correlation suffices, and there is little need to distinguish between causation and confounding. Off-the-shelf statistical and machine-learning models are good enough, because they exploit correlations between input variables and the target variable.

Delineating causation from correlation is essential for many other applications. If our task is to predict the expression of Y in a cell where we inhibit the expression of X (Figure 1g), such as by gene silencing, it is essential to know which of the models below is the correct one: $X \rightarrow Y$, or $Y \rightarrow X$, or $X \leftarrow U \rightarrow Y$. Because if $X \rightarrow Y$ is true, expression of Y will become residual; given the other two models, expression of Y will not be affected. If we do not have a model, collecting more data does not help us with the task.

The simple example illustrates three critical points: (i) causality is required to predict the outcome of intervention; (ii) data alone are not able to tell causality, no matter how much data we have and however complex or fancy the algorithm is; and (iii) we need both scientific models, that is, prior knowledge and hypothesis, and data to infer causality.

The takeaway message is that although statistical and machine-learning models are useful if we ask for correlation, we need additional tools to recover causality from correlations if our goal is to intervene in the system. While correlation helps us to predict or even to explain,[15,16] causality helps us to understand and to act.

Causal modelling with directed acyclic graphs (DAGs)

Directed acyclic graphs, or DAGs, are used as computational models of causal inference. DAGs contain nodes, which represent variables, and edges, which represent causal relationships. Like the two-variable cases described above, DAGs represent either knowledge or hypotheses about causal relationships between the variables. The edge, that is, the causal relationship, can have any discrete, linear, or nonlinear functional form. A lack of edge between any pair of nodes means that we exclude the possibility that they have direct causality on each other. An indirect causal relationship may still exist in such cases due to the propagation of causality through the DAG, as we shall see below.

Three three-node structures are prevalent in DAGs. Understanding these common structures, known as *motifs*, allows us to interpret more complex causal models:

1. The pipe (Figure 2a) describes the simplest chain of causality: $X \rightarrow Z \rightarrow Y$.
2. The fork (Figure 2b) describes a variable Z causing both X and Y: $X \leftarrow Z \rightarrow Y$.
3. The collider (Figure 2c) describes the situation where a variable Z is caused by both X and Y: $X \rightarrow Z \leftarrow Y$.

A pipe transduces causality in a chain, such as *Vemurafenib binding to V600E-mutated BRAF → reduced BRAF signalling → reduced tumour size*. If the intermediate cause is manipulated, such as by mutations enhancing BRAF signalling, the causal effect from drug on tumour size will diminish or even reverse. Statistically speaking, X and Y are marginally correlated, and they are independent conditional on Z.

A fork, like the temperature example raised before, generates correlations between variables that do not cause each other, for example: *PI3K phosphorylation ← RAS signalling → BRAF proliferation*. If the common cause is present, both its effects are correlated; if the common cause is absent, or if we focus on the subset of the data where the common cause takes the same value, the effects become independent. Similar to the situation of pipes, X and Y are marginally correlated, and they become independent conditional on Z.

Colliders behave very differently from pipes and forks. For instance, cells that lose the expression of the BOP1 (block of proliferation 1) gene become resistant against vemurafenib.[17] The mechanism can be described by a collider structure: *BRAF V600E → increased phosphorylation of ERK1/2 ← loss of BOP1*, namely both BRAF V600E mutation (V600E for short) and loss of BOP1 (BOP1↓) increases ERK1/2 phosphorylation (↑pERK1/2). As long as either condition is satisfied, we can observe ↑pERK1/2. Assuming that they are the only two factors influencing pERK1/2, the collider creates a dependency between V600E and BOP1↓. When ↑pERK1/2 is true, either V600E or BOP1↓ or both must exist; otherwise, neither V600E nor BOP1↓ is possible. Even if V600E and BOP1↓ are statistically independent from each other, that is, knowing V600E does not give us any information about whether BOP1 is lost or not, the two events become correlated once we observe ↑pERK1. In contrast to pipes and forks, X and Y are marginally independent but become correlated conditional on Z.

Knowledge of causal structure allows us to better recognize cognitive bias. We demonstrate this point with the example of colliders. With simulated data generated by the three three-node structures in Figure 1, we show that we can remove or

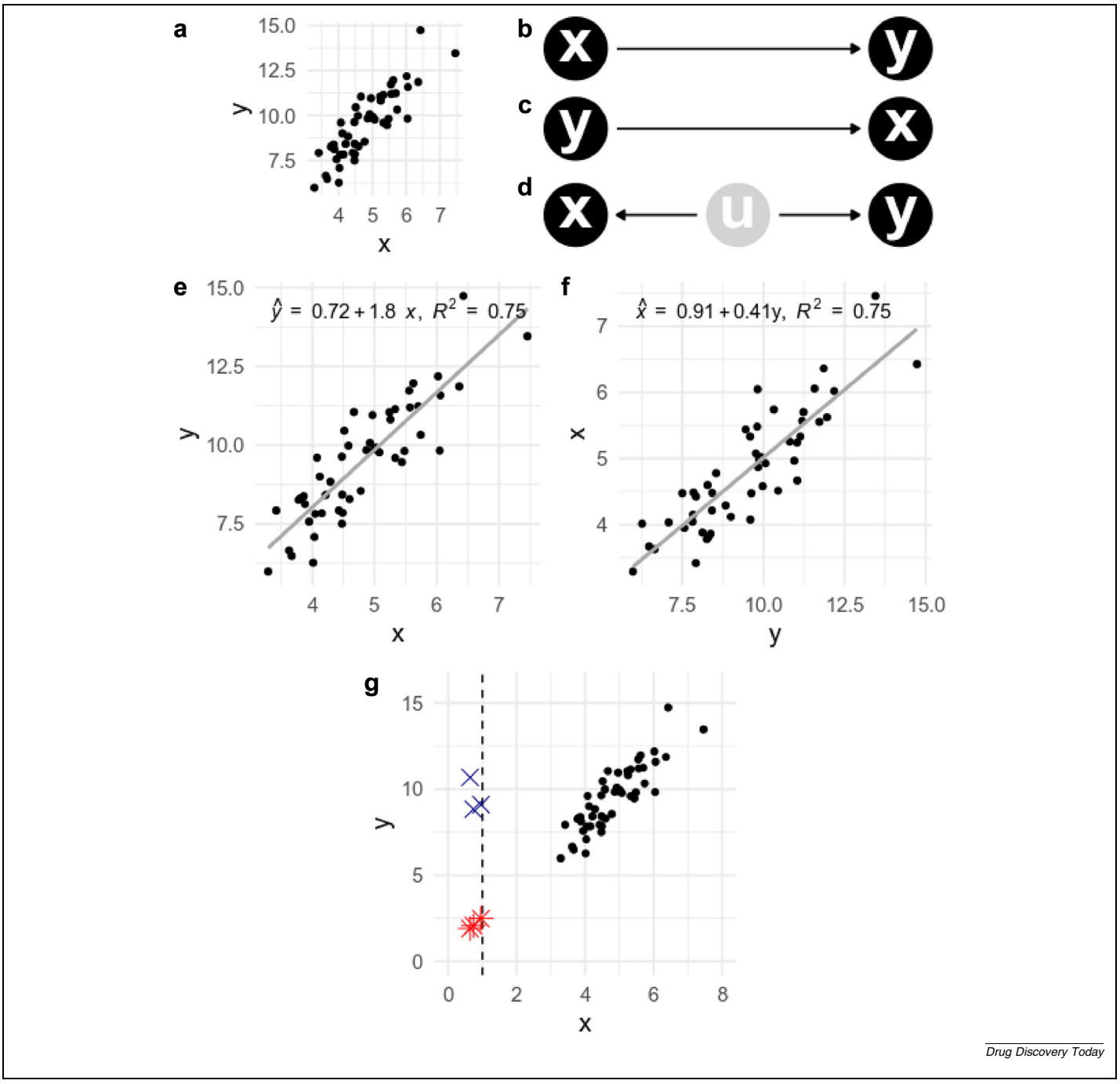

FIGURE 1

Distinguishing between correlation and causation. (a) Correlation between two variables *x* and *y*, which represents the expression levels of two proteins X and Y in a population of cells, respectively. Each dot represents one cell. The data were simulated by a linear relationship. (b-d) Three causal models that can generate the correlation observed in (a). two direct causal relationships (panel b and c) and one with a confounding variable (panel d). Coincidence, conspiracy, collider, and chronology as reasons generating correlations are not shown for simplicity. (e-f) Data and statistical models alone cannot tell the direction of causation, or whether confounding variables exist. Panel e shows the linear regression with *x* as independent variable and *y* as dependent variable using data shown in panel A. Panel f shows the regression with the same dataset, with the role of *x* and *y* swapped. The correlation coefficients (R^2) of both regression models are identical. (g) Predicting the outcome of interventions requires causal models. Recall that we use *x* and *y* to denote expression of two proteins X and Y in a population of cells. And now we reduce the expression of X artificially to 1.0 (dashed line in black). Depending on the causal structure, the outcome of Y may be near to the red stars (model in panel b) or blue crosses (models in panel c or d).

create correlation between *x* and *y* if we stratify by the third variable *z*, depending on the graph structure. The conclusion holds if *z* is a continuous variable and if we regress it out in a statistical model. Imagine now we face a dataset with two variables, *x* and *y*; we observe a good linear correlation; and our goal is to predict the value of *y*, given a new set of values *x*. Does a linear model suffice? Not necessarily, because if there is a variable *z* forming a collider with *x and* y, and its value changes for the

new set of values x, then our linear model is likely to fail. Even the simplest causal graphs show us it is important to know which factors affect the generation of data.

Besides the pipes, forks, and colliders, we highlight two four-node motifs:

1. The descendant describes a variant of the collider, where the variable caused by two variables X and Y is not directly observed, but its descendant, namely another variable that is caused by the unobserved variable, is observed. Because we often use the variable U to indicate an unobserved variable, the DAG can be represented as $X \rightarrow U \leftarrow Y$ and $U \rightarrow Z$. The structure, for instance, is useful to model the effect of known risk factors (X) and treatment (Y) on disease status (U). Though U may not be directly observed, a biomarker (Z), which is partially affected by U, can be measured.
2. The paw or 3-pan describes a class of motifs, including the one shown in Figure 5a. The tail variable is known as the instrumental variable. Models with instrumental variables are applicable in a wide variety of contexts, including Mendelian randomization, a technique for target identification to be described in Section 4, and the analysis of clinical trial data with noncompliance, that is, when some patients in the treatment arm do not take the medicine as prescribed.[18]

DAGs consisting of these and other motifs[19] can represent causal relationships between any number of variables. We refer readers interested in reading and understanding DAGs better to several outstanding tutorials.[20–22]

Because DAG structure specifies marginal and conditional correlations, it is possible to rank models by their plausibility, given data, and to learn the strength of causal relationships. The two tasks are known as *causal identification*, or *causal discovery*, in which the truth value of a claim of the form 'C causes E' is determined, and quantitative *causal estimation*, in which a numerical value s (strength) is estimated for a claim of the form 'C has an effect on E.' Causal discovery is more challenging than causal estimation, and discovered causal effects usually require causal estimation before they can be used for high-stakes applications.[23,24] Nevertheless, it plays an important role in reconstructing causal networks for disease understanding and target identification, especially because we lack much biological knowledge on how the observed data are generated. See Section 4 for details.

Relationship between causal inference and established modelling techniques

Causal inference completes statistical modelling and mechanistic modelling, two commonly used techniques in drug discovery and development (Table 1). On one hand, statistical modelling, including pattern recognition–based machine-learning models, sometimes termed *artificial intelligence* (AI), aims at identifying prediction-relevant features and functions that transform the features to approximate the target variable and thereby exploit correlations. The models may or may not have a causal interpretation. On the other hand, mechanistic modelling uses mathematical models of biological processes to describe and predict how components, information, or energy of a system evolve with time. Examples include pharmacokinetics and pharmacodynamics models and quantitative system pharmacology models based on ordinary or partial differential equations, agent-based models, and hidden Markov models. Although mechanistic models usually have a causal interpretation and are powerful for predicting the outcome of an intervention, we often miss information of important factors that affect the system's behaviour.

The three modelling approaches complement and benefit from each other. Reciprocal benefits between statistical and causal models are apparent. By discovering and quantifying causal relationships between observed variables, we can render statistical models' causal interpretations. By using advanced machine-learning approaches, we may learn highly complex functional forms describing the causal effect[25] or discover causality from high-dimensional data.[26] Practically, because we are often limited by the volume of high-quality data, causal inference may likely profit from parsimonious statistical models and Bayesian approaches.[27]

Results of causal inference help refine and improve existing mechanistic models by including causal factors and removing confounding factors. In turn, outputs from mechanistic models complement the current causal inference regime by predicting time-dependent outcomes of intervention, in particular those of components that regulate each other.

Causal inference for experimental data and observational data

The power of causal inference to connect and enhance existing modelling approaches is particularly prominent in the analysis of observational studies. Classically, we distinguish observational studies from controlled experimental studies. In a controlled experiment, we assign test objects, such as animals in preclinical experiments, to groups. One group receives the treatment and the other group does not. If we go one step further to require that the grouping is randomised with regard to any relevant attributes of the test objects (passage, sex, body weight, etc.), then we have a randomised controlled experiment, a gold standard to establish causality. In an observational study, in contrast, we measure or survey members of a sample without trying to affect them. Such data can come from epidemiological studies, electronic health records (EHRs), insurance claims, and other data that come from natural (instead of controlled) experiments, such as omics and behavioural data of healthy individuals and patients.

Several reasons make it imperative to use causal models to analyse data generated by observational studies. First, it allows us to integrate knowledge and hypotheses about inevitable biases in the data generation process. Second, we can investigate the causal effect of independent variables of interest while considering other variables that affect the outcome, known as *covariates*. Third, we can potentially resolve the effects of variables that influence both the independent variable and the outcome, known as *confounding variables*. If we analyse observational data as if we were handling randomised experimental data, without considering the causal structure underlying the data generation and collection, we may derive false, sometimes ridiculous, conclusions.[28]

Even if a study is set up as a classical randomised experiment, various reasons may break the randomization and demand causal

TABLE 1

A comparison of data modelling techniques.

	Examples	Prediction for independent and identically distributed samples	Prediction for outcome of intervention	Answering *what if*, counterfactual questions	Data-driven discovery
Mechanistic models	Pharmacokinetics and physically based pharmacokinetic models	Yes	Yes	Yes	Maybe
Causal models	See discussion and examples	Yes	Yes	Yes	Maybe
Statistical models	Cox regression model for survival, statistical tests, (generalised) linear models for *omic* data, support vector machines, random forest, neural networks	Yes	No	No	Yes

Adapted from Table 1.1 in Peters *et al.*[137].

inference to identify the treatment effect. For instance, in clinical trials, noncompliance is a common issue, where some patients do not take the treatment as prescribed. To identify the real treatment effect while acknowledging the noncompliance, it is necessary to employ a causal model, with the *instrumental variable* model illustrated above as the simplest example. Other issues that have been successfully addressed by causal inference include missing data,[29,30] such as when patients drop out of the trial, and intercurrent events, that is, events occurring after randomization that can either preclude observation of the outcome of interest or affect its interpretation, as defined by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use in its Guideline E9(E1), such as the development of antidrug antibodies.[31–34] The book by Guido Imbens and Donald Rubin provides an outstanding introduction to the analysis of these and other aberrations from classical randomised experiments with causal inference.[18]

Causal inference can leverage both experimental and observational data to generate insight. Eichler *et al.*[35] proposed a new framework, known as *threshold crossing*, that generates evidence by synthesising historical randomised clinical trials and real-world data resources. A recent landscape assessment and elicited comments have also shown that causal inference penetrates and contributes significantly to clinical trial study design and analysis by judging the suitability and integrating real-world data from observational studies.[36,37]

Finally, we emphasise the importance of study design for both experimental and observational studies. It is widely acknowledged that well-designed randomised controlled experiments, both in preclinical drug discovery[38] and in clinical drug development,[39] are the gold standard of verifying causal relationships. The quality and strength of evidence provided by an observational study and causal inference is also determined largely by the study design.[40] Study design should be scrutinised prior to causal inference in order to gauge whether we can answer the question of interest.

A step-by-step guide to causal inference

We introduce causal inference as an iterative process of six steps: *modelling* → *identification* → *estimation* → *refutation* → *refinement* → *application*. They are illustrated in Figure 3 and detailed below. To assist practitioners acquiring hands-on experience with causal inference, we offer complementary interactive tutorials Supplementary material. The tutorials, implemented in the programming languages Python (with *Jupyter Notebooks*) and R (with *Rmarkdown)*, are available at https://github.com/Accio/causal_-drug_discovery.

Step 1: Modelling

To start, we construct a DAG to model causal mechanisms that generated the observed data by synthesising common sense, scientific knowledge, or explicit assumptions. If multiple hypotheses exist, we may build several DAGs and subject them to analysis and comparison.

In case there is little knowledge and hypothesis available, we may apply causal discovery techniques to raise hypotheses about the mechanism of data generation. Commonly used methods include Bayesian networks,[41] factor analysis,[42] and encoder-decoder and other deep generative models.[43] Limitations of such methods include (i) they are often technically challenging (*NP-hard* problems), (ii) the results often include many alternative hypotheses that explain the data equally well, and (iii) the proposed models may not be causal, that is, they fail to be validated by interventional studies. Nevertheless, the discovered causal graphs, combined with expert review and curation, may serve as reasonable starting points.

Step 2: Identification

Once a causal model is set up, we test whether we can answer the question of interest quantitatively. The quantity that addresses the causal question, such as *how strong is the effect of the drug on disease progression*, is known as the *estimand* (see more details in Supplementary File 1, in the Supplementary material online). In this step, our goal is to assess whether we can estimate the estimand from the data at all, because some graph structures prohibit us from doing so.

In a causal model, the estimand is usually the causal effect pointing from a treatment (cause) to its target (effect) variable, which can be a regression coefficient or other numeric values that quantify the strength of the causal relationship. The causal graph model structure determines which estimators are available, and the numerical values of estimates are gauged by the users to interpret the results and challenged by refutation analysis.

An interesting and important result from theoretical studies is that we can assert whether an estimand can be identified using

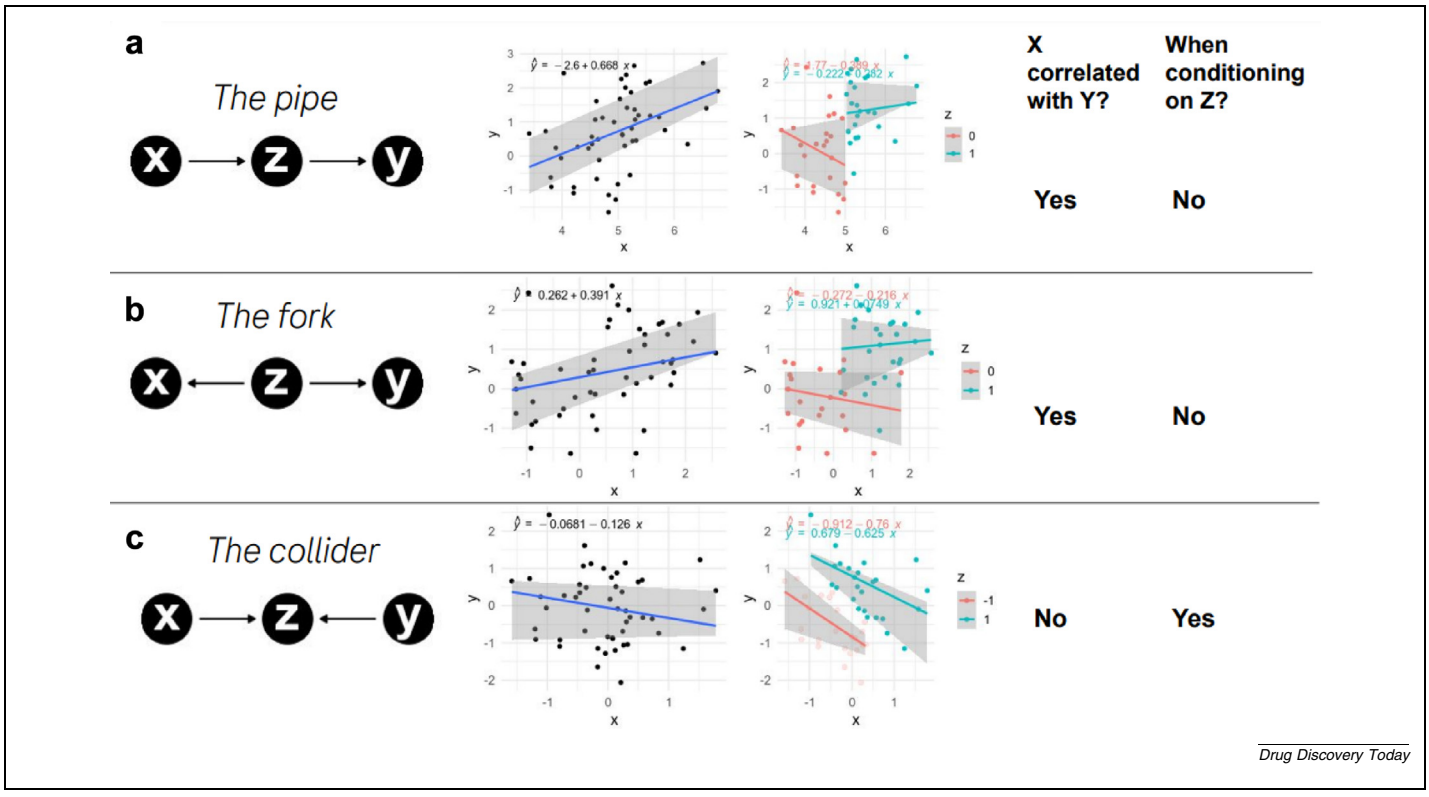

FIGURE 2

Common 3-node structures in DAGs. (a) Left. the pipe structure consists of a chain of three variables. The model was used to generate a simulated dataset of 50 data points. We specified that *x* follows a Gaussian distribution with mean of 5 and standard deviation, *z* takes the value of 0 if $x < 5$ and 1 otherwise, and *y* follows another Gaussian distribution with mean defined by 2*z and standard deviation. Middle. simulated data visualised with scatter plot. Each dot represents one data point. Both plots show *x* on the X-axis and *y* on the Y-axis. The positions of the points are the same in both plots. In the left plot, the regression line (blue) and its confidence interval (grey) are shown for $y \sim x$; in the right plot, the regression lines and confidence intervals are shown dependent on the value of Z. A simple visual assistance is that if *x* and *y* are marginally (mid-left plot) or conditionally (mid-right plot) correlated with each other, then the confidence interval should *not* contain any horizontal grid line. In this case, it is clear that while *x* is correlated with *y*, the correlation is broken if we condition on *z*. This summary is shown on the right panel of the plot. (b) Similar to panel a, with the fork structure on the left. Simulation rules (data points N = 50). *z* follows a Bernoulli distribution with a probability of success of 0.5. Both *x* and *y* follow normal distribution with mean defined by *z* and standard deviation. The interpretation of middle and right panels is comparable to panel a. (c) Similar to panel a and b, with the collider structure on the left. Simulation rules (data points N = 50). Both *x* and *y* follow normal distribution with 0 mean and standard deviation. The value of *z* is 1 if x + y > 0 and −1 otherwise. The interpretation of middle and right panels is comparable to panel a and b. The code to generate both the simulated data and the visualisations is available at https://github.com/Accio/causal_drug_discovery/blob/main/2021-12-CausalSalad.Rmd.

graphical models alone, independent of the choice of the estimator and without access to the data. This means that we can tell whether the question can be answered at all by causal inference *before* we collect or generate any data, which is usually the most time- and labour-consuming step in knowledge acquisition.

In this step, we need to identify the target estimand by specifying the quantity of interest in the DAG model. Usually, we use software to identify whether the estimand can be estimated based on the graph structure. If the estimand is not identifiable, we have little choice other than modifying our model, redefining our question, or admitting that the problem cannot be solved with causal inference. Otherwise, the model is said to be *identifiable*, and we can continue with the estimation step.

A causal model is intrinsically a generative model, namely it can be used to generate simulated data. Once the identifiability is established, a common practice is to simulate data with the model with specified parameters, to run the inference with simulated data, and to examine both whether the simulated data reassemble observed data and whether the estimate is consistent with the input parameters. Inference with simulated data offers us experience with the model, opportunities to identify bugs and flaws of the model before the real run, and confidence in the model and the choice of estimator.

Step 3: Estimation

Next, we collect existing data or generate new data according to the DAG. The importance of quality control of data cannot be overestimated: *garbage in, garbage out*. Data may be filtered so that only the subset that satisfies the DAG structure is used for the estimation.

Once we have decided on the estimand and the data, we can estimate the causal effect by identifying an appropriate estimator and deriving the estimate from the data, which is usually done by software (see Box 2 for popular open-source software packages for causal inference).

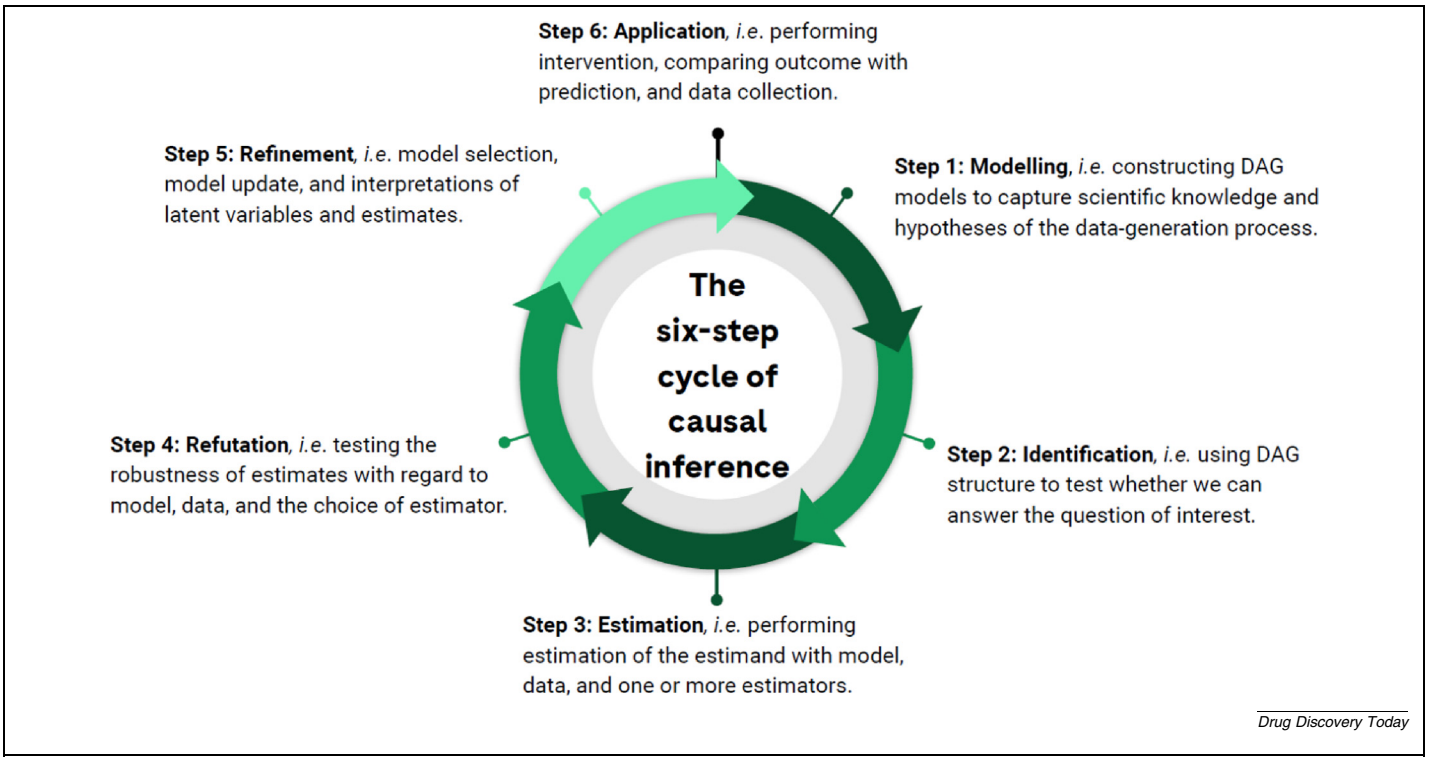

FIGURE 3
A six-step model of causal inference, detailed in Section *A step-by-step guide of causal inference*.

BOX 2 Emerging software tools for causal inference.

- https://github.com/microsoft/dowhy[44]
- Dagitty[150] in R (GNU), https://www.dagitty.net/
- https://github.com/babylonhealth/counterfactual-diagnosis
- https://github.com/lingfeiwang/findr
- https://github.com/causal-machine-learning
- https://github.com/uber/causalml
- https://github.com/microsoft/econml
- https://github.com/google/CausalImpact
- https://github.com/IBM/causallib
- https://github.com/uhlerlab/causaldag
- https://github.com/tlverse/tlverse, with tutorial at https://tlverse.org/tlverse-handbook/.

The choice of estimator depends on the DAG structure and the nature of the data. The popular *DoWhy* package, for instance, offers diverse estimators depending on the graph structure, including (i) (generalised) linear regression; (ii) stratification, matching, or weighting by propensity score, namely the probability of each unit being assigned to a treatment group given a set of observed covariates; (iii) using instrumental variables; and (iv) machine-learning–based estimators implemented in the *EconML* package. Although the choice of appropriate estimators is context- and question-specific, several measures can help us make better choices, especially (i) using simulated data, (ii) learning from case studies and past experience, and (iii) applying refutation techniques, discussed in the next step.

Step 4: Refutation

Although estimation is commonly deemed the most important step in causal inference, in high-stakes applications such as drug discovery and development, we need to test the robustness of our estimates. This is achieved by *refutation*, a collective name of many modelling techniques to test the strength, validity, and our confidence of the estimated causal effect. Commonly used techniques are listed in Supplementary File 1, in the Supplementary material online.

In addition, one can perform data partition or bootstrapping to estimate the variability of the estimates. Furthermore, bespoke refutation techniques can be used to address application-specific questions, such as noncompliance in randomised clinical trials.[44] We refer interested readers to a recent review about these and other techniques known as *sensitivity analysis*.[45]

Statistical causal inference is not the only way to identify causality (Box 1). Alternative evidence from other models, such as results from mechanistic models and results from randomised controlled experiments, should be used to strengthen or challenge our belief in the output of the causal model.

BOX 1 An anarchy of approaches to causality.

Despite a long history of causality research in multiple disciplines, especially statistics, computer science, epidemiology, healthcare, and philosophy of science, there is no consensus way of defining and inferring causality.[28] Impactful approaches for causal inference include the counterfactual outcome framework, the Campbell's framework, Bayesian decision theory, and the directed acyclic graph (DAG) approach.[142] We note that they are not mutually exclusive. In this review, we particularly focus on the counterfactual framework and the DAG approach.

A variety of philosophical and scientific approaches have been developed to investigate causality. Below is a list of approaches that mostly impacted the authors:

- The constant conjunction by David Hume.
- Path analysis pioneered by Sewall Wright in 1920s.[143]
- Potential outcomes proposed by Jerzy Neyman in 1923, which revealed the possibility of achieving causal inference with randomised trials.[144]
- Hill's causality criteria in 1960s.[2(p19)]
- Ignorability proposed by Donald Rubin in 1970s, known as unconfoundedness to epidemiologists and selection on observables by economists, which lead to propensity scores.[18,145,146]
- Bayesian network, DAGs, and do operators proposed between 1980s and 2000s by Judea Pearl.[83,147]
- Representation learning for causal inference.[132]
- Causal machine learning, aiming at individualised causal inference using high-dimensional features to subdivide population, with emerging work in medicine and healthcare.[47,148]
- Learning causal relationships, representing a more substantial challenge than causal effect learning.[24]
- Challenges of estimation from finite sample, see Shipley.[149]

We close this box with a quote from Richard McElreath: 'There is no method for making causal models other than science. There is no method to science other than honest anarchy.'

Step 5: Refinement

Even when the estimated causal effects withstand refutation analysis, they are seldom the sole goal of our investigation. Given that we lack a comprehensive understanding for virtually all biological problems, we often need to *refine* the model, which includes selection from multiple models, addition of new variables, interpretation of latent variables and of estimates, and answering *what if* questions. The refined model can be used, on one hand, to guide design of new experiments to further refute or update the model and, on the other hand, to guide interventions.

Step 6: Application

One of the ultimate goals of causal inference is to perform interventions in the real world. The outcomes can be further analysed in the causal inference framework, therefore closing the loop.

In contrast to a 'hypothesis-free' paradigm, where one expects to learn mechanisms generating the data from data alone, the cycle of *modelling, identification, estimation, refutation, refinement,* and *application* integrates knowledge, hypothesis, and data to address scientific questions. Interdisciplinary teamwork is required to construct explicit models of causality, to collect estimand-specific data, and to identify causal relationships.

Literature review and case studies

Literature review

To gain an overview of applications of causal inference in drug discovery and development, we compiled a list of publications by querying the MEDLINE/PubMed database. We found more than 800 scientific publications (Table S1 in the Supplementary material online) up to September 2022. As a comparison, we also queried publications on machine learning and artificial intelligence and stratified the number of publications by years (Figure 4a). We note two interesting patterns: (i) publications on causal inference in drug discovery have increased almost steadily since 1990, and (ii) the topic receives much less coverage than machine learning and artificial intelligence. Despite the scope of the latter two concepts being admittedly much broader, the patterns underscore the importance of populating the causal mindset and language among practitioners.

We classified the literature by applications and observed that causal inference has been applied throughout the value chain of drug discovery and development (Figure 4b,c). The majority of publications are about analysis of observational studies, especially drug safety, pharmacovigilance, and real-world data. Methodological papers and publications reporting applications for target identification and assessment, clinical trial design and analysis, and biomarker studies follow with distance. Applications in other areas, in particular preclinical and translational research, are still scarce.

Given the broad application of causal inference, and given that applications for clinical data, real-world data, and observational studies have been extensively reviewed elsewhere,[36,37,46,47] we focus below on case studies of causal inference for disease understanding and target identification in translational research.

Translational research includes activities that aim at establishing the causal relationship between drug candidates and disease progression. It has two interlinked components: forward and reverse translation. Forward translation predicts drug dosing and risk/benefit ratio in patients with *in silico* models and data from *in vitro*, *ex vivo*, and *in vivo* animal models and microdosing human studies. Reverse translation informs and improves decision-making in forward translation by analysing data collected in clinical trials, real-world data, as well as data generated with clinical stage or marketed molecules in preclinical models. Forward and reverse translation form a closed loop and complement each other to validate or refute the proposed causal relationship between treatment and patient health status.

Although translational research has many facets, we choose disease understanding and target identification to highlight the impact of causal inference for two main reasons. First, leveraging the causal engine in the target identification and selection phase may have the strongest potential in reducing costs to society of disease and of drug discovery.[48–50] Second, applications in this area involve heterogeneous and high-dimensional data, such as

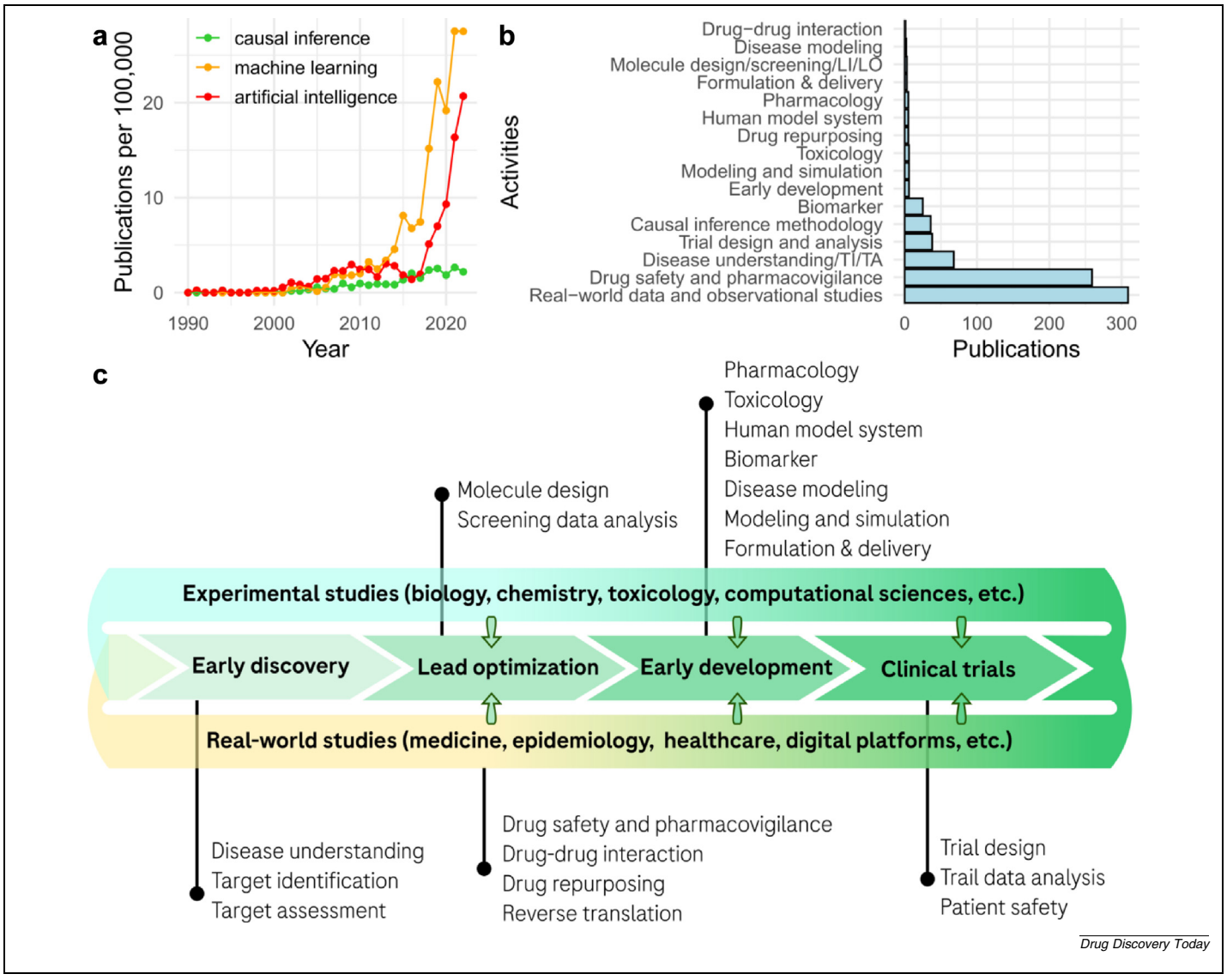

FIGURE 4

Established and emerging applications of causal inference along the value chain of drug discovery and development. (a) Publications indexed by MEDLINE/PubMed with keywords *causal inference* (green), *machine learning* (yellow), and *artificial intelligence* (red). Only peer-reviewed papers relevant for drug discovery and development are included. Numbers of publications are divided by the total number of publications and multiplied by 100,000. (b) Classification of publications on causal inference in drug discovery and development into activity categories. We read the abstract, and if available, the full text, of identified causal-inference publications and manually categorised them. The activities are reversely ordered by the number of total publications. (c) We positioned activities found in the literature review in the context of both forward and reverse translation. In the middle we show a simplified diagram of forward translation consisting of four steps. early discovery, lead optimization, early development, and clinical trials. Above and below we show reverse translation, i.e. analysis of experimental and observational real-world data, and its feedback to forward translation. Applications of causal inference are highlighted besides the lollipops.

EHRs and genomic and other omic data. Although the data type varies in other tasks of translational research, such as risk/benefit assessment, biomarker selection, and patient stratification and enrichment, the concepts, software tools, and practice can be transferred.

Learning causal associations from natural experiments and observational studies

Although target identification was driven by druggability and *in vitro* or animal disease models for a long time, reverse translation is upending this pattern. Causal inference contributes evidence to support or disaffirm drug targets by analysing data from natural experiments and observational studies. The need for causal inference is particularly strong in emerging fields such as microbiome medicine, where high-quality data are relatively scarce and associative analyses are still prevalent.[51] Similarly, causal inference is needed when the study of disease aetiology is challenging, such as in neuroscience. Path analysis, a precursor to and variant of causal inference, has been applied to prioritise or refute targets in neurogenesis, Parkinson's disease, and autism spectrum disorder.[52,53,54,55] See Supplementary File 1 (in the Supplementary material online) for a detailed case study with Alzheimer's disease.

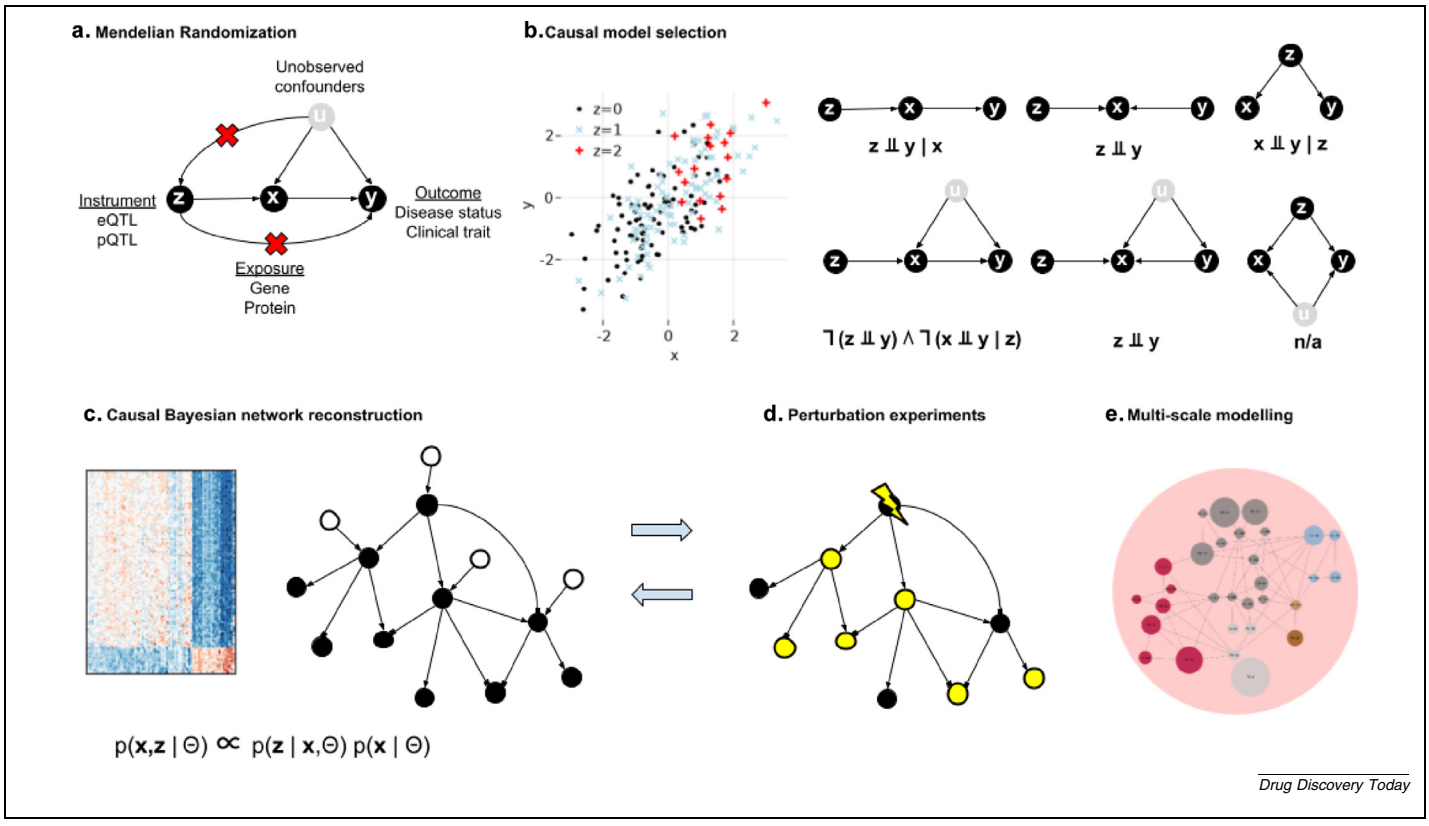

FIGURE 5

Causal inference for disease understanding and target identification. (a) Mendelian Randomization (MR) estimates the causal effect of an exposure ("x", e.g. a gene or protein) on an outcome ("y", e.g. disease status or clinical trait) using a genetic instrument for the exposure ("z", an eQTL or pQTL for a gene or protein exposure). MR is applicable if the instrument is independent of unobserved confounders ("u") and if the outcome is associated with the instrument only through the exposure (no alternative paths) (indicated by red crosses). (b) Simulated scatter plot of two coexpressed genes X and Y, with samples coloured according to the genotype of a genetic marker Z for X. Model selection determines the causal model underlying the data by testing the conditional independencies implied by each causal DAG. Representative causal models without (top) and with (bottom) unobserved confounders are shown. Model selection assumes that the edge Z → X must be included in the model and that Z is independent of any unobserved confounders. (c) When a larger set of coexpressed genes is considered (left), the data is modelled by a causal Bayesian network (right). Black nodes represent genes (x) and white nodes genetic instruments (z) that are used to orient the causal directions of gene-gene edges, and the Bayesian network represents the probability distribution of jointly observing *x* and z given a set of model parameters Θ learned from the training data. (d) Controlled perturbation experiments give direct information about the variables causally downstream of the perturbed node (yellow nodes) and can be used to refute or refine networks reconstructed from observational data. (e) Models inferred at one scale (e.g. cell-type or tissue-specific gene or protein networks) can be integrated into higher-level models. In the figure, nodes represent tissue-specific causal Bayesian networks (node colour, tissue; node size, network size) and edges represent "eigengene" similarities. Panel E obtained from Talukdar, Husain A., et al. Cell systems 2.3 (2016). 196–208 under CC-BY-NC-ND licence.

Application of causal inference on multiomic data (genomics, transcriptomics, proteomics, metabolomics) leads us to identify causal genes, pathways, and gene regulatory networks that have a direct effect on disease states.[56–59] Genome-wide association studies (GWASs) have mapped the genetic architecture of common diseases in humans[60] and have identified genetic loci that affect drug response and susceptibility to adverse drug reactions.[61] Drugs targeting genetically supported targets were reported to be twice as likely to be successful as other drugs.[62,63] Exploiting GWAS data in drug discovery is challenging, however, because there are generally hundreds to thousands of genetic risk variants, mostly lying in noncoding genomic regions and each typically contributing only a small amount of risk.

Mendelian randomization (MR)[64,65] is a statistical approach that uses genetic variants as randomised instruments ('natural experiments') to identify causal associations between heritable traits. It is based on the fact that genotypes are independently assorted and randomly distributed in a population by Mendel's laws, or in a less strict manner, independent of each other outside haplotype blocks,[66] and not affected by environmental or genetic confounders that affect both the exposure (target manipulation) and the outcome (disease state). If it can be assumed that a genetic locus affects the outcome only through the exposure, then the causal effect of the exposure on the outcome can be derived from their relative associations to the genetic locus, which acts as an instrumental variable[67] (Figure 5a).

By integrating GWAS with studies that map genetic effects on the transcriptome or proteome (expression or protein quantitative trait loci; eQTLs or pQTLs, respectively), MR can identify causal associations between molecular traits and disease.[68] When applied to proteomic data, inferred causal associations between proteins and disease can suggest drug repurposing opportunities[69] or new candidate drug targets.[70,71] An advantage of MR is that it can be performed using summary statistics alone.[72] A lim-

itation is that MR tests effects of molecular traits on diseases one by one and that no molecular pathways are reconstructed. This is important because the candidate causal factors may act indirectly or redundantly with other factors, or the candidate factors may not be druggable themselves but affect intermediate druggable targets.

When genetics and multiomic data from individuals in a segregating population are merged, comparison of independent, causative, and reactive DAGs can orient the direction of causality among QTLs and pairs of correlated genes or proteins[73,74] (Figure 5b). Statistical model selection compares the conditional independence relationships implied by each DAG and uses maximum likelihood to select the model that best fits the available data. Statistical significance is estimated by expressing the conditional independence implications of the tested DAGs as combinations of likelihood ratio tests. Either their maximum *P*-value is treated as an omnibus hypothesis test[75–77], or false discovery estimation is used to express the results of the individual tests as probabilities of their null or alternative hypothesis being true.[78,79] Accounting for hidden confounders requires additional assumptions on the genetic instrument, usually satisfied if the instrument is a *cis*-acting eQTL or pQTL for the 'exposure' variable, and involves a trade-off between a high false-negative rate or an increased false-positive rate.[79,80]

Bayesian networks combine the results of pairwise causal inferences into a causal network model of the underlying biological system. Bayesian networks are probabilistic machine-learning models for expressing both prior knowledge and inferred conditional independence and causal relationships among variables.[81–83] A Bayesian gene network consists of a DAG which connects regulatory genes to their targets[84,85] (Figure 5c). The structure of a Bayesian network is inferred from the data using score-based or constraint-based methods,[82] where the results of pairwise causal inference tests are used to constrain the DAG search space.[84–90] This is done by expressing the joint likelihood of observing both data types as a standard likelihood term for observing gene expression data, given the DAG, and an additional likelihood term for observing the genotype data, given the expression data and causal interactions implied by the DAG[84,85,88–90] (Figure 5c).

To bridge from the molecular scale to disease states, clinical phenotypes are not usually modelled as nodes in the causal network. Instead, genome-wide omic data are first partitioned into coexpression modules associated with clinical traits,[91–94] and causal networks are learned for each module separately and connected in a higher-level Bayesian or coexpression network where each node is a module 'eigengene,' a representation of the module's expression profile by its first principal component[95] (Figure 5e). This higher-level network can model tissue-specific and intertissue communication processes that are the result of the 'collective' states of molecular networks and that are more proximal to physiological disease processes.[58,93,94,96]

As in almost all other areas of genomics, single-cell technologies are creating new opportunities for learning causal disease mechanisms from population-based data.[97] Cost-effective strategies for generating single-cell RNA-sequencing data across individuals have been developed,[98,99] and the first large-scale population genetics studies have been published recently.[100–104] A particularly attractive application will be to exploit the fact that allele-specific expression can be mapped in single cells.[105] This will allow us to quantify eQTL effect sizes[106] at the level of an individual instead of a population, which will lead to a better understanding of the variation of causal effects across individuals.

Learning causal associations from controlled perturbation experiments

Having considered the impact of causal inference on reverse translation, now we turn to its impact on forward translation by analysing controlled perturbation experiments (Figure 5d).

For target-based discovery programs, canonical gene knockout or perturbation experiments in model systems establish causality in a manner similar to randomised controlled trials. When these experiments are combined with genome-wide readouts, they provide insights into the direct and indirect causal targets of the perturbed gene. Systematic small-molecule and perturbation screens using RNAi technology followed by transcriptome sequencing have been performed in cancer cell lines.[107] CRISPR-based technology combined with single-cell RNA sequencing is now rapidly expanding the scope of conducting genome-wide perturbation screens in human cells.[108–112]

Causal inference from perturbation experiments allows us to disentangle direct from indirect effects and predict the outcome of a new intervention or perturbation. Nested effect models are probabilistic graphical models to infer a genetic hierarchy from the nested structure of observed perturbation effects.[113,114] A more formal approach uses causal DAGs and *do*-calculus to model causal effects from perturbation screens.[115,116]

Pioneers and experts in phenotypic drug discovery have long realised the importance of establishing causal relationships. Moffat *et al.*[117] proposed a model of *chain of translatability*, namely from the chemical matter to the assay phenotype, to the preclinical disease model, and finally to human disease. We note the multiscale nature of this chain of translatability and the resemblance of challenges faced by target-based programs and those faced by phenotypic programs, except for the target identification and assessment question. A key open question is whether causal inference, and in particular causal discovery techniques, can help phenotypic programs identify molecular targets and mode of action.

A fundamental issue with forward translation is that cellular or animal models differ from human physiology. Therefore, a critical additional request is that the identified link must be persistent in humans in the form of *in vitro–in vivo* translation, governed by physiological constraints, and between-species translation, governed by evolutionary forces.

In order to decide whether experimental findings can be extrapolated across domains that differ both in their distributions and in their inherent causal characteristics, we need prior scientific knowledge about the invariance of certain mechanisms, represented in 'selection diagrams,' DAGs in which the causal mechanisms are explicitly encoded and in which differences in populations are represented as local modifications of those mechanisms.[118,119] Switching between populations corresponds to conditioning on different values of the variables that locate the mechanisms where structural discrepancies between

the two populations are suspected to take place. The transportability problem can then be formulated and decided using *do*-calculus and graphical criteria on the selection diagrams.

Translational study is undergoing a profound change. Although animal models are traditionally used as a bridge between *in vitro* experiments and studies with humans, the field is witnessing strong interest and investment in human model systems such as stem cell–derived cell lines, primary cells, organoids, organ-on-a-chip, or other microphysiological systems. They are currently being tested for many purposes, especially efficacy and safety assessment, prediction and modelling of pharmacokinetic and pharmacodynamic parameters, biomarker identification, and disease modelling.[120]

From the point of view of causal inference, the key question is how much the causal model relevant for the disease and the drug candidate in human systems is conserved in model systems, let it be *in vitro*, *in vivo*, or human model systems. To answer this question, it is necessary to establish causal models with each model system and to compare the models between the systems. So far, the translational value of most human model systems is judged by *looking alike*, that is, to assess similarity in morphology and/or omic profiles with regard to primary human material. In order to mimic the causal relationship, more work is required to assess these systems by *functioning alike*, that is, to assess their response upon perturbation and conditions that mimic disease aetiology and development.[121,122]

Discussion

This review addresses three questions: (i) When do we need a causal model? (ii) How can causal inference be performed? (iii) Which activities in drug discovery have been or will be empowered by causal inference? With the 6C model of correlation and causation in mind, we argued that causal models are essential for predictive modelling of the outcome of intervention, for answering *what if* questions, and for understanding disease biology and why drugs work (or do not). We introduced causal inference as a six-step iterative process with hands-on examples in the complementary tutorial. Finally, we performed an extensive literature review and discussed recent application of causal inference in translational research.

As discussed in the introduction, causality is multiscale. This renders causal inference a suitable tool to investigate interactions between drugs and human biological systems, which are *per se* multiscale.[123] Leveraging advances in statistical modelling and mechanism-based multiscale modelling has the potential to transform heterogeneous data into unified knowledge, which helps us understand the impact of genetic, epigenetic, and environmental factors on drug action.[124]

An emerging consensus is that the quality of *in silico* model predictions outweighs the speed of prediction, and qualitative statements that enable scientists to focus on promising targets and regions of chemical space are often more impactful than quantitative predictions for decision-making.[125,126] In line with this, we foresee three key opportunities for causal inference: (i) high-quality prediction for predictive modelling of interventional outcomes, including out-of-distribution predictions[125,126]; (ii) individualised causal inference and estimation of heterogeneous treatment effects[127]; and (iii) causally and counterfactually based decision-making.[20,128]

Besides the opportunities, we foresee three major challenges to be overcome by researchers and practitioners of causal inference in the coming decade: (i) the methodology of causal inference needs to be further developed to embrace the reality and complexity of biology; (ii) the community needs to share and have open access to high-quality data and models; and (iii) we need a change in both language and mindset.

The methodology of causal inference warrants further research. DAGs, though very powerful, have intrinsic limitations when used to model biological systems. Edges in DAGs must be acyclic; that is, we are never trapped in a closed loop by following the directions of edges. DAGs are therefore not suited for analysing reciprocal causal relationships, though they are prevalent in biology across scales from molecular interactions[129] to consciousness.[130] To model systems with such relationships, one can either take advantage of longitudinal data to identify causal relationships[131], or model the net output of nodes involved in reciprocal relationships as a variable instead of modelling individual nodes, or employ computational models other than DAGs, such as graph neural networks and other representation learning techniques.[132] Both the theory and the software required for performing such analyses need further research and development. Despite recent progress discussed above,[26] it is still challenging to perform causal inference with high-dimensional data, for systems with multiple complex traits, that is, many factors with small effect sizes,[133–135] and for complex adaptive systems where feedback loops abound.[136–138]

Open contribution and access to improve biological models is essential for the community to refine and leverage causal models. Resources of causal models are emerging quickly,[139] and community-wide efforts such as DREAM challenges have been hosted to identify causal relationships from *omic* datasets.[140] Nevertheless, the causal language remains foreign to many researchers; we do not always distinguish correlation from causation explicitly, and, in many cases, scientific findings are not represented as causal models. Although it remains a complex and unsolved problem of how to populate the causal language and how to encourage researchers to share data and causal models, we are cautiously optimistic about further development because of the potential gain. The Linus's Law formulated by Eric S. Raymond,[141] *Given enough eyeballs, all bugs are shallow,* is a motivation for all of us to adopt the causal language and share the causal models: To share and to refine causal models in a community is to understand how biology and drugs work collectively.

Several reasons may make practitioners of drug discovery and development, including many colleagues we have interviewed, wary of causal inference. We need the front-up payment of bringing up knowledge and hypothesis, often exposing our naivety and lack of knowledge. In a real-world setting, setting up a model is particularly challenging because we often have incomplete information and inaccurate knowledge about human biology and the pharmacology and toxicology of drugs, not to mention the abundance of invisible or unquantifiable quantities and events that affect the efficacy and safety of drugs. We have to give up the hope about 'learning from data alone' and 'hypothesis-free approaches.' We know that machine-learning

models such as deep neural networks and graphical neural networks can learn patterns if they are trained with enough data. Isn't it better to accumulate more data and let the machine find out the causality? Why do we stick to and propose causal inference?

We welcome these questions and doubts. They lead us to believe that we need a change in both language and mindset to think and talk about causality. We argue that (i) causality exists beyond data, and it is not possible to learn causality from data alone without prior knowledge and hypotheses; and (ii) the central task of drug discovery and development is to discover and validate multiscale causal relationships. The true causal network, in contrast to causal DAG models, can contain feedback loops and therefore cannot be subject to the analysis introduced here as a whole. However, the network can be separated into smaller subnetworks, each verified by the six-step cycle of causal inference, and the causality propagates between the scales. The network connects the drug molecule with disease outcome via its interactions with biological molecules. The interactions emerge as pharmacology and toxicology on cellular, organ-and-system levels. Furthermore, the network can be enriched with individual traits such as genetics, medical history, lifestyle, and environmental factors to allow population analysis.

Concluding remarks

Causal inference offers a principled approach to model- and data-driven predictive modelling and decision-making. It complements statistical and mechanistic models to empower an understandable synthesis and integration of knowledge and data across scales. We foresee that further development of its methodology, open-accessible causal models and high-quality data, and a broader adoption of the causal language and mindset will fuel both forward and reverse translation. In short, causal inference empowers us to seek and to prove causality with new drugs.

Funding

T.M. is supported by grants from the Research Council of Norway (grant numbers 312045 and 331725). J.D.Z.'s work is funded by F. Hoffmann-La Roche Ltd.

Declarations of interest

No interests are declared.

Data availability

We have shared the data as supplementary information.

Acknowledgements

Both authors thank colleagues, students, and team members for discussion, input, and feedback for the manuscript. J.D.Z. wishes to explicitly thank input from the members of the Predictive Modelling and Data Analytics (PMDA) chapter, especially Zhiwen Jiang, Milad Adibi, Juliane Siebourg-Polster, Tony Kam-Thong, Balazs Banfai, Sarah Morillo Leonardo, Martin Ebeling, Cheikh Diack, Anja Lieberherr, Benjamin Ribba, and Julia Pletz, as well as visitors of the weekly *Bioinformatics Club*. We also thank Kaspar Rufibach, Denis Engemann, Gonzalo Christian Duran-Pacheco, and Nicoleta Spinu for sharing their insights, criticism, and feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.drudis.2023.103737.

References

1. Beebee H, Hitchcock C, Menzies PC, eds. *The Oxford Handbook of Causation*. Oxford University Press; 2009.
2. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295–300.
3. Williamson J. Establishing causal claims in medicine. *Int Stud Philos Sci*. 2019;32:33–61. https://doi.org/10.1080/02698595.2019.1630927.
4. Stewart AK. How thalidomide works against cancer. *Science*. 2014;343:256–257. https://doi.org/10.1126/science.1249543.
5. Vargesson N. Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Res*. 2015;105:140–156. https://doi.org/10.1002/bdrc.21096.
6. Ito T et al. Identification of a primary target of thalidomide teratogenicity. *Science*. 2010;327:1345–1350. https://doi.org/10.1126/science.1177319.
7. Békés M, Langley DR, Crews CM. PROTAC targeted protein degraders: the past is prologue. *Nat Rev Drug Discov*. 2022;21:181–200. https://doi.org/10.1038/s41573-021-00371-6.
8. Deshaies RJ. Multispecific drugs herald a new era of biopharmaceutical innovation. *Nature*. 2020;580:329–338. https://doi.org/10.1038/s41586-020-2168-1.
9. Tummino TA et al. Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science*. 2021;373:541–547. https://doi.org/10.1126/science.abi4708.
10. Fischer H, Kansy M, Bur D. CAFCA: a novel tool for the calculation of amphiphilic properties of charged drug molecules Accessed 23 October 2019.. *Chim Int J Chem*. 2000 https://www.ingentaconnect.com/content/scs/chimia/2000/00000054/00000011/art00006.
11. Berkson J. Limitations of the application of fourfold table analysis to hospital data (reprint). *Int J Epidemiol*. 2014;43:511–515. https://doi.org/10.1093/ije/dyu022.
12. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51–63. https://doi.org/10.1016/0021-9681(79)90012-2.
13. Reichenbach H, Reichenbach M. *The Direction of Time*. Dover; 1999.
14. Hitchcock C, Rédei M. Reichenbach's common cause principle Accessed 31 July 2023.. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University; 2021.
15. Shmueli G. To explain or to predict? *Stat Sci*. 2010;25:289–310. https://doi.org/10.1214/10-STS330.
16. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206–215. https://doi.org/10.1038/s42256-019-0048-x.
17. Gupta R, Bugide S, Wang B, Green MR, Johnson DB, Wajapeyee N. Loss of BOP1 confers resistance to BRAF kinase inhibitors in melanoma by activating MAP kinase pathway. *Proc Natl Acad Sci*. 2019;116:4583–4591. https://doi.org/10.1073/pnas.1821889116.
18. Imbens GW, Rubin DB. *Causal inference for statistics, social, and biomedical sciences: an introduction*. Cambridge University Press; 2015.
19. Cinelli C, Forney A, Pearl J. *A crash course in good and bad controls*. *Soc Sci Res Netw*. 2020. https://doi.org/10.2139/ssrn.3689437.
20. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Method*. 2008;8:70. https://doi.org/10.1186/1471-2288-8-70.
21. Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams: pitfalls and tips. *J Epidemiol*. 2020;30:153–162. https://doi.org/10.2188/jea.JE20190192.
22. Laubach ZM, Murray EJ, Hoke KL, Safran RJ, Perng W. A biologist's guide to model selection and causal inference. *Proc R Soc B Biol Sci*. 1943;2021:20202815. https://doi.org/10.1098/rspb.2020.2815.

23. Spirtes P, Zhang K. Causal discovery and inference: concepts and recent methodological advances. *Appl Inform*. 2016;3:3. https://doi.org/10.1186/s40535-016-0018-x.
24. Guo R, Cheng L, Li J, Hahn PR, Liu H. A survey of learning causality with data: problems and methods. *ACM Comput Surv*. 2020;53:75:1–75:37. https://doi.org/10.1145/3397269.
25. Luo Y, Peng J, Ma J. When causal inference meets deep learning. *Nat Mach Intell*. 2020;2:426–427. https://doi.org/10.1038/s42256-020-0218-x.
26. Lopez R, Hütter JC, Pritchard JK, Regev A. *Large-Scale Differentiable Causal Discovery of Factor Graphs*. Published online 15 June 2022. https://doi.org/10.48550/arXiv.2206.07824.
27. Vasudevan RK, Ziatdinov M, Vlcek L, Kalinin SV. Off-the-shelf deep learning is not enough, and requires parsimony, Bayesianity, and causality. *npj Comput Mater*. 2021;7:1–6. https://doi.org/10.1038/s41524-020-00487-0.
28. McElreath R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. CRC Press: Taylor and Francis; 2020.
29. National Research Council (U.S.), National Research Council (U.S.), National Academies Press (U.S.), eds. *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press; 2010.
30. Little RJ. Missing data assumptions. *Annu Rev Stat Appl*. 2021;8:89–107. https://doi.org/10.1146/annurev-statistics-040720-031104.
31. Bornkamp B, Bermann G. Estimating the treatment effect in a subgroup defined by an early post-baseline biomarker measurement in randomized clinical trials with time-to-event endpoint. *Stat Biopharm Res*. 2020;12:19–28. https://doi.org/10.1080/19466315.2019.1575280.
32. Bornkamp B et al. Principal stratum strategy: potential role in drug development. *Pharm Stat*. 2021;20:737–751. https://doi.org/10.1002/pst.2104.
33. Heinzmann D, Kong S. Principal stratum strategy to investigate anti-drug antibody impact on outcome in randomized controlled trials. Presented at: https://bbs.ceb-institute.org/wp-content/uploads/2020/09/06-Heinzmann.pdf
34. Kong S, Heinzmann D, Lauer S, Lu T. Weighted approach for estimating effects in principal strata with missing data for a categorical post-baseline variable in randomized controlled trials. *ArXiv210104263 Stat*. Published online January 11, 2021. Accessed 18 November 2021. https://arxiv.org/abs/2101.04263
35. Eichler H et al. 'Threshold-crossing': a useful way to establish the counterfactual in clinical trials? *Clin Pharmacol Ther*. 2016;100:699–712. https://doi.org/10.1002/cpt.515.
36. Ho M et al. The current landscape in biostatistics of real-world data and evidence: causal inference frameworks for study design and analysis. *Stat Biopharm Res*. 2021;15:43–56. https://doi.org/10.1080/19466315.2021.1883475.
37. Hampson LV, Degtyarev E, Tang (Sammi) R, Lin J, Rufibach K, Zheng C. Comment on 'Biostatistical considerations when using RWD and RWE in clinical studies for regulatory purposes: a landscape assessment'. *Stat Biopharm Res*. 2021;15:23–26. https://doi.org/10.1080/19466315.2021.1994459.
38. Lazic SE. *Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility*. Cambridge University Press; 2016.
39. Zhang W, Ramezani R, Naeim A. Causal Inference in medicine and in health policy, a summary. *arXiv:210504655v4*. Published online 10 May 2021. https://doi.org/10.48550/arXiv.2105.04655
40. Rosenbaum PR. *Design of Observational Studies*. Springer; 2010.
41. Scutari M, Denis JB. *Bayesian Networks: With Examples in R*. 2nd ed. CRC Press; 2021.
42. Fabrigar LR, Wegener DT. *Exploratory Factor Analysis*. Oxford University Press; 2012.
43. Lopez R, Gayoso A, Yosef N. Enhancing scientific discoveries in molecular biology with deep generative models. *Mol Syst Biol*. 2020;16. https://doi.org/10.15252/msb.20199198 e9198.
44. Sharma A, Kiciman E. DoWhy: An end-to-end library for causal inference. *ArXiv201104216 Cs Econ Stat*. Published online 9 November 2020. Accessed 29 September 2021. https://arxiv.org/abs/2011.04216
45. Razavi S et al. The future of sensitivity analysis: an essential discipline for systems modeling and policy support. *Environ Model Softw*. 2021;137. https://doi.org/10.1016/j.envsoft.2020.104954 104954.
46. Lipkovich I, Ratitch B, Mallinckrodt CH. Causal inference and estimands in clinical trials. *Stat Biopharm Res*. 2020;12:54–67. https://doi.org/10.1080/19466315.2019.1697739.
47. Prosperi M et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat Mach Intell*. 2020;2:369–375. https://doi.org/10.1038/s42256-020-0197-y.
48. Plenge RM. Disciplined approach to drug discovery and early development. *Sci Transl Med*. 2016;8. https://doi.org/10.1126/scitranslmed.aaf2608 349ps15.
49. Plenge RM. Priority index for human genetics and drug discovery. *Nat Genet*. 2019;51:1073–1075. https://doi.org/10.1038/s41588-019-0460-5.
50. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12:581–594. https://doi.org/10.1038/nrd4051.
51. Lv BM, Quan Y, Zhang HY. Causal inference in microbiome medicine: principles and applications. *Trends Microbiol*. 2021. https://doi.org/10.1016/j.tim.2021.03.015.
52. Lazic SE. Using causal models to distinguish between neurogenesis-dependent and -independent effects on behaviour. *J R Soc Interface*. 2012;9:907–917. https://doi.org/10.1098/rsif.2011.0510.
53. Gündner AL et al. Path mediation analysis reveals GBA impacts Lewy body disease status by increasing α-synuclein levels. *Neurobiol Dis*. 2019;121:205–213. https://doi.org/10.1016/j.nbd.2018.09.015.
54. Yap CX et al. Autism-related dietary preferences mediate autism-gut microbiome associations. *Cell*. 2021. https://doi.org/10.1016/j.cell.2021.10.015.
55. Tosun D et al. Contribution of Alzheimer's biomarkers and risk factors to cognitive impairment and decline across the Alzheimer's disease continuum. *Alzheimers Dement*. 2021;18:1370–1382. https://doi.org/10.1002/alz.12480.
56. Rockman MV. Reverse engineering the genotype–phenotype map with natural genetic variation. *Nature*. 2008;456:738–744.
57. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. *Nat Rev Drug Disc*. 2009;8:286–295.
58. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009;461:218–223.
59. Li Y, Tesson BM, Churchill GA, Jansen RC. Critical reasoning on causal inference in genome-wide linkage and association studies. *Trends Genet*. 2010;26:493–498.
60. Uffelmann E et al. Genome-wide association studies. *Nat Rev Methods Primer*. 2021;1:59. https://doi.org/10.1038/s43586-021-00056-9.
61. Daly AK. Genome-wide association studies in pharmacogenomics. *Nat Rev Genet*. 2010;11:241–246. https://doi.org/10.1038/nrg2751.
62. Nelson MR et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–860. https://doi.org/10.1038/ng.3314.
63. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet*. 2019;15:e1008489.
64. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1–22.
65. Evans DM, Davey SG. Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu Rev Genomics Hum Genet*. 2015;16:327–350. https://doi.org/10.1146/annurev-genom-090314-050016.
66. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9:477–485. https://doi.org/10.1038/nrg2361.
67. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res*. 2007;16:309–330.
68. Zhu Z et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48:481–487. https://doi.org/10.1038/ng.3538.
69. Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. *Nat Rev Genet*. 2021;22:658–671. https://doi.org/10.1038/s41576-021-00387-z.
70. Bretherick AD et al. Linking protein to phenotype with Mendelian Randomization detects 38 proteins with causal roles in human diseases and traits. *PLoS Genet*. 2020;16:e1008785.
71. Pietzner M et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021;374. https://doi.org/10.1126/science.abj1541 eabj1541.
72. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658–665. https://doi.org/10.1002/gepi.21758.
73. Jansen RC, Nap JP. Genetical genomics: the added value from segregation. *Trends Genet*. 2001;17:388–391.
74. Schadt EE et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet*. 2005;37:710–717.
75. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet*. 2009;10:23.
76. Millstein J, Chen GK, Breton CV. cit: hypothesis testing software for mediation analysis in genomic applications. *Bioinformatics*. 2016;32:2364–2365.

77. Badsha M, Fu AQ. Learning causal biological networks with the principle of Mendelian randomization. *Front Genet*. 2019;10:460.
78. Chen LS, Emmert-Streib F, Storey JD. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol*. 2007;8:R219.
79. Wang L, Michoel T. Efficient and accurate causal inference with hidden confounders from genome-transcriptome variation data. *PLoS Comput Biol*. 2017;13 e1005703.
80. Ludl AA, Michoel T. Comparison between instrumental variable and mediation-based methods for reconstructing causal gene networks in yeast. *Mol Omics*. 2021;17:241–251.
81. Friedman N. Inferring cellular networks using probabilistic graphical models. *Science*. 2004;308:799–805.
82. Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press; 2009.
83. Pearl J. *Causality: Models, Reasoning and Inference*. 2nd ed. Cambridge University Press; 2009.
84. Neto EC, Keller MP, Attie AD, Yandell BS. Causal graphical models in systems genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. *Ann Appl Stat*. 2010;4:320.
85. Hageman RS, Leduc MS, Korstanje R, Paigen B, Churchill GA. A Bayesian framework for inference of the genotype–phenotype map for segregating populations. *Genetics*. 2011;187:1163–1170.
86. Zhu J et al. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res*. 2004;105:363–374.
87. Zhu J et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet*. 2008;40:854–861.
88. Tasaki S, Sauerwine B, Hoff B, Toyoshiba H, Gaiteri C, Neto EC. Bayesian network reconstruction using systems genetics data: comparison of MCMC methods. *Genetics*. 2015;199:973–989.
89. Wang L, Audenaert P, Michoel T. High-dimensional Bayesian network inference from systems genetics data using genetic node ordering. *Front Genet*. 2019;10:1196. https://doi.org/10.3389/fgene.2019.01196.
90. Scutari M, Howell P, Balding DJ, Mackay I. Multiple quantitative trait analysis using Bayesian networks. *Genetics*. 2014;198:129–137.
91. Zhang B et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013;153:707–720.
92. Beckmann ND et al. Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. *Nat Commun*. 2020;11:3942. https://doi.org/10.1038/s41467-020-17405-z.
93. Talukdar H et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell Syst*. 2016;2:196–208.
94. Koplev S et al. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat Cardiovasc Res*. 2022;1:85–100. https://doi.org/10.1038/s44161-021-00009-1.
95. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol*. 2007;1:54.
96. Seldin MM et al. A strategy for discovery of endocrine interactions with application to whole-body metabolism. *Cell Metab*. 2018;27:1138–1155.
97. van der Wijst MGP, de Vries DH, Brugge H, Westra HJ, Franke L. An integrative approach for building personalized gene regulatory networks for precision medicine. *Genome Med*. 2018;10:96. https://doi.org/10.1186/s13073-018-0608-4.
98. Kang HM et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol*. 2018;36:89–94. https://doi.org/10.1038/nbt.4042.
99. Mandric I et al. Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. *Nat Commun*. 2020;11:5504. https://doi.org/10.1038/s41467-020-19365-w.
100. Bryois J et al. Cell-type specific cis-eQTLs in eight brain cell-types identifies novel risk genes for human brain disorders. *Nat Neurosci*. 2022;25:1104–1112.
101. LifeLines Cohort Study et al. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet*. 2018;50:493–497. https://doi.org/10.1038/s41588-018-0089-9.
102. Neavin D et al. Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol*. 2021;22:76. https://doi.org/10.1186/s13059-021-02293-3.
103. Yazar S et al. Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. *Science*. 2022;376. https://doi.org/10.1126/science.abf3041 eabf3041.
104. Nathan A et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature*. 2022;606:120–128. https://doi.org/10.1038/s41586-022-04713-1.
105. Jiang Y, Zhang NR, Li M. SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol*. 2017;18:74. https://doi.org/10.1186/s13059-017-1200-8.
106. Mohammadi P, Castel SE, Brown AA, Lappalainen T. Quantifying the regulatory effect size of *cis*-acting genetic variation using allelic fold change. *Genome Res*. 2017;27:1872–1884. https://doi.org/10.1101/gr.216747.116.
107. Subramanian A et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171:1437–1452.
108. Dixit A et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*. 2016;167:1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038.
109. Adamson B et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*. 2016;167:1867–1882.e21. https://doi.org/10.1016/j.cell.2016.11.048.
110. Jaitin DA et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell*. 2016;167:1883–1896.e15. https://doi.org/10.1016/j.cell.2016.11.039.
111. Datlinger P et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*. 2017;14:297–301. https://doi.org/10.1038/nmeth.4177.
112. Schraivogel D et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat Methods*. 2020;17:629–635. https://doi.org/10.1038/s41592-020-0837-5.
113. Markowetz F, Kostka D, Troyanskaya OG, Spang R. Nested effects models for high-dimensional phenotyping screens. *Bioinformatics*. 2007;23:i305–i312. https://doi.org/10.1093/bioinformatics/btm178.
114. Tresch A, Markowetz F. Structure learning in nested effects models. *Stat Appl Genet Mol Biol*. 2008;7. https://doi.org/10.2202/1544-6115.1332.
115. Meinshausen N, Hauser A, Mooij JM, Peters J, Versteeg P, Bühlmann P. Methods for causal inference from gene perturbation experiments and validation. *Proc Natl Acad Sci*. 2016;113:7361–7368. https://doi.org/10.1073/pnas.1510493113.
116. Wang Y, Solus L, Yang KD, Uhler C. Permutation-based causal inference algorithms with interventions. *ArXiv170510220 Stat*. Published online 4 November 2017. Accessed 16 December 2021. https://arxiv.org/abs/1705.10220
117. Moffat JG. Turning the light on in the phenotypic drug discovery black box. *Cell Chem Biol*. 2017;24:545–547. https://doi.org/10.1016/j.chembiol.2017.05.005.
118. Pearl J, Bareinboim E. External validity: from Do-Calculus to transportability across populations. *Stat Sci*. 2014;29. https://doi.org/10.1214/14-STS486.
119. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci*. 2016;113:7345–7352. https://doi.org/10.1073/pnas.1510507113.
120. Ewart L, Roth A. Opportunities and challenges with microphysiological systems: a pharma end-user perspective. *Nat Rev Drug Discov*. 2021;20:327–328. https://doi.org/10.1038/d41573-020-00030-2.
121. Lukonin I et al. Phenotypic landscape of intestinal organoid regeneration. *Nature*. 2020;586:275–280. https://doi.org/10.1038/s41586-020-2776-9.
122. Fleck JS et al. Inferring and perturbing cell fate regulomes in human cerebral organoids 2021:2021.08.24.457460. *bioRxiv*. 2021. https://doi.org/10.1101/2021.08.24.457460.
123. Zhang JD, Sach-Peltason L, Kramer C, Wang K, Ebeling M. Multiscale modelling of drug mechanism and safety. *Drug Discov Today*. 2020;25:519–534. https://doi.org/10.1016/j.drudis.2019.12.009.
124. Xie L et al. Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput Biol*. 2014;10:e1003554.
125. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. *Drug Discov Today*. 2020. https://doi.org/10.1016/j.drudis.2020.12.009.
126. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. *Drug Discov Today*. 2021;26:1040–1052. https://doi.org/10.1016/j.drudis.2020.11.037.
127. Bica I, Alaa AM, Lambert C, van der Schaar M. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin Pharmacol Ther*. 2021;109:87–100. https://doi.org/10.1002/cpt.1907.

128. Jung J, Concannon C, Shroff R, Goel S, Goldstein DG. Simple rules for complex decisions. *ArXiv170204690 Stat*. Published online April 2, 2017. Accessed 26 November 2019. https://arxiv.org/abs/1702.04690
129. Alon U. *An introduction to systems biology: design principles of biological circuits*. Chapman & Hall/CRC; 2007.
130. Bartels A. Consciousness: what is the role of prefrontal cortex? *Curr Biol*. 2021;31:R853–R856. https://doi.org/10.1016/j.cub.2021.05.012.
131. Shojaie A, Fox EB. Granger causality: a review and recent advances. *Annu Rev Stat Appl*. 2022;9:289–319. https://doi.org/10.1146/annurev-statistics-040120-010930.
132. Schölkopf B, et al. Towards causal representation learning. Published online 22 February 2021. Accessed 5 August 2021. https://arxiv.org/abs/2102.11107v1
133. de Boer CG, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, Regev A. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat Biotechnol*. 2020;38:56–65. https://doi.org/10.1038/s41587-019-0315-8.
134. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.
135. Liu X, Li YI, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. *Cell*. 2019;177:1022–1034.e6. https://doi.org/10.1016/j.cell.2019.04.014.
136. Peters J, Bauer S, Pfister N. Causal models for dynamical systems. *ArXiv200106208 Math Stat*. Published online 17 January 2020. Accessed 18 November 2021. https://arxiv.org/abs/2001.06208
137. Peters J, Janzing D, Schölkopf B. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press; 2017.
138. Holland JH. Complex adaptive systems. *Daedalus*. 1992;121:17–30.
139. Status of causality in biological databases: data resources and data retrieval possibilities to support logical modeling. Brief Bioinform 2021;22:bbaa390. https://academic.oup.com/bib/article/22/4/bbaa390/6055722
140. Cell-to-cell and type-to-type heterogeneity of signaling networks: insights from the crowd. Mol Syst Biol 2021;17(10):e10402. https://doi.org/10.15252/msb.202110402
141. Raymond ES. *The cathedral & the bazaar: musings on linux and open source by an accidental revolutionary*. 1st ed. O'Reilly; 1999.
142. Thoemmes F. Ausgewählte Kausalitätstheorien im Vergleich. *Gesundheitswesen*. 2011;73:880–883.
143. Wright S. The relative importance of heredity and environment in determining the piebald pattern of Guinea-Pigs. *Proc Natl Acad Sci*. 1920;6:320–332. https://doi.org/10.1073/pnas.6.6.320.
144. Splawa-Neyman J, Dabrowska DM, Speed TP. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Stat Sci*. 1990;5:465–472. https://doi.org/10.1214/ss/1177012031.
145. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91:444–455. https://doi.org/10.1080/01621459.1996.10476902.
146. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66:688–701. https://doi.org/10.1037/h0037350.
147. Pearl J, Glymour M, Jewell NP. *Causal inference in statistics: a primer*. Wiley; 2016.
148. Pearl J. The seven tools of causal inference, with reflections on machine learning. *Commun ACM*. 2019;62:54–60. https://doi.org/10.1145/3241036.
149. Shipley B. *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference with R*. Cambridge University Press; 2016.
150. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol*. 2016;45:1887–1894. https://doi.org/10.1093/ije/dyw341.