We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Introduction: An Overview of AI in Oncology Drug Discovery and Development

Kristofer Linton-Reid

Abstract

Artificial intelligence (AI) has been termed the machine for the fourth industrial revolution. One of the main challenges in drug discovery and development is the time and costs required to sustain the drug development pipeline. It is estimated to cost over 2.6 billion USD and take over a decade to develop cancer therapeutics. This is primarily due to the high numbers of candidate drugs failing at late drug development stages. Many sizable pharmaceutical and biotech companies have made considerable investments in AI. This is primarily due to recent advancements in AI, which have displayed the possibility of rapid low-cost drug discovery and development. This overview provides a general introduction to AI in drug discovery and development. This chapter will describe the conventional oncology drug discovery pipeline and its associated challenges. Fundamental AI concepts are also introduced, alongside historical and modern advancements within AI and drug discovery and development. Lastly, the future potential and challenges of AI in oncology are discussed.

Keywords: oncology, AI, drug discovery, drug development

1. Introduction

Artificial intelligence (AI) has been termed the machine for the fourth industrial revolution. AI is anticipated to transform every industry. In drug discovery and development, the key challenges are the time and costs required to sustain the drug development pipeline. It is estimated to cost over 2.6 billion USD and take over a decade to develop an oncology therapeutic [1]. These soaring costs are mostly a result of money invested in the 90% of candidate therapies that fail at the late stages of drug development, between phase 1 trials and regulatory approval [2]. AI is projected to be the foundation for an era of quicker, cheaper, and more efficient drug discovery and development.

Recent advancements in AI are displaying the possibility of rapid low-cost drug discovery and development. The term AI broadly describes the ability of a machine to perform tasks commonly associated with intelligent beings. Another term, machine learning (ML), is a subset of AI involving machines using data to artificially think for themselves. The main difference between ML and AI is that ML is the direct application and involves the combination and analysis of complex, disparate data sets.

Within the pharmaceutical industry, experts agree that AI will revolutionize and change how drugs are discovered. There are many components, directly and indirectly, related to the drug discovery and development that AI can enhance. These include but are not limited to: the use of AI in tumour classification [3], computer-aided organic synthesis [4], compound discovery [5], assay development, and biomarker and target discovery [6–8]. In general, AI aims to automate and optimize slow processes to substantially speed up the R&D drug discovery process.

Several pharmaceutical, biotech, and software companies are also making every effort to integrate AI with drug discovery and development. In 2016, Pfizer partnered with IBM Watson Health, an AI platform, to enhance their search for immuno-oncology treatments. Sanofi paired with Dundee university spin-out Exscientia, to discover metabolic-disease therapies. In 2009, Roche acquired Genentech for $46.8 billion, providing a foundation for Roche's biotechnology division, which is not integrating AI. Genentech is now collaborating with GNS Healthcare platform to use machine learning to find and validate potential new drug candidates. Recently, Genentech displayed the capacity of AI to diagnose diabetic macular degeneration.

Even large traditional tech companies are investing in drug development. Alphabet's subsidiary DeepMind developed an AI platform, AlphaFold, that predicted protein 3D structures based upon genomic data; their prediction was better than over 90 other companies including Novartis, and Pfizer, in the 13th Critical Assessment of Structure Prediction. DeepMind's success with AlphaFold is displaying how non-healthcare companies can also contribute to and improve the drug discovery and development pipeline. These investments are forming a clear vision that AI will play an important role in future drug discovery and development.

In this overview, we start with introducing key components of conventional oncology drug discovery, and associated shortfalls. Following this, fundamental AI concepts are introduced, alongside historical and modern advancements within AI and drug discovery and development. Lastly, the future potential and challenges of AI in oncology are introduced.

2. Conventional oncology drug discovery and development

The conventional drug discovery and development pipeline has five key components: target identification, lead discovery, preclinical development, clinical development, and regulatory approval (**Figure 1**). A drug discovery program initiates after researching the inhibition or activation of a protein or pathway and explaining the potential therapeutic effect. This leads to the selection of a biological target, often requiring extensive validation prior to the lead drug discovery phase. This phase involves the search for a viable drug-like small molecule or biological

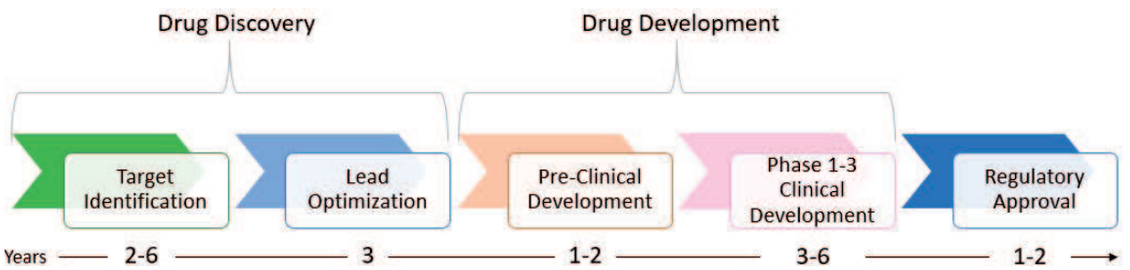


Figure 1.
Five key components of drug discovery and development: target identification, lead discovery, preclinical development, clinical development, and regulatory approval.

therapy, termed a development candidate. The drug candidate will progress into preclinical development, and if successful into clinical development.

2.1 Drug discovery and development pipeline: target identification and validation

Biological target identification and validation is a fundamental step in drug discovery. A biological target is a broad term, used to describe a variety of entities including proteins, metabolites, and genes. A biological target must have a clear effect, meet clinical and therapeutic needs, as well as industry needs. Above all, a biological target must be 'druggable'. The term 'druggable' refers to a target that can be bound by a small molecule or larger biologic and elicit a response.

2.1.1 Target identification

A variety of methods exist to identify biological targets. This includes gene expression, proteomics and genomics analysis, and phenotypic screening.

The analysis of mRNA/protein expression is often employed to elucidate expression to disease relationships if changes in expression levels are correlated with exacerbation or progression. At the genetic level, targets are identified by determining if there is an association between genetic polymorphism and disease occurrence or progression. For example, one of the most well-studied genetic-disease associations is that of *N*-acetyltransferase 2 (*NAT2*) with bladder and colon cancer. *N*-acetyltransferase 1 (*NAT1)* and *NAT2* are precursors of enzymes that mediate the transformation of two types of carcinogens, aromatic and heterocyclic amines. The *NAT2* rapid acetylator phenotype and the slowest NAT2 acetylator phenotype are associated with colon and bladder cancer respectively [9, 10].

Phenotypic screening is another method for target identification. This can take a variety of forms. Generally, compounds are screened in cellular or animal disease models to identify a compound that leads to the desired change in phenotype. Kurosawa and colleagues [11] screened for overexpressed carcinoma antigens by isolating human monoclonal antibodies that bind to the surface of tumour cells. In this study, clones were screened with immunostaining. Clones that displayed strong staining with the malignant cells were selected. Subsequently, 21 distinct antigens were derived via mass spectroscopy. Several immunotherapies may be capable of binding to these 21 antigen targets, possibly leading to a new clinical therapy.

Target identification may involve one or a combination of the previously mentioned methods.

2.1.2 Target validation

While identifying a target typically requires one method, the following target validation requires a variety of methods. A multi-validation approach increases confidence in the biological target and subsequent drug candidate's success.

There are a variety of target validation methods that may be implemented, although validation almost always requires target expression in the disease-relevant cells or tissues. A typical primary validation protocol is to measure the expression of protein and/or mRNA in clinical samples, with immunohistochemistry and in situ hybridization.

Generally, in vivo studies are often a pivotal factor in the decision to proceed with drug development; these usually involve protein inhibition/gene knock-out/knock-in studies. Transgenic animal models are particularly useful as they facilitate phenotypic observations. These animal models often yield insights into potential

therapeutic side effects. Transgenic models traditionally gene edits whereby an animal would lack or obtain a certain gene(s) for its entire life. An example is the P2X7 knockout mouse model, which lacks an inflammatory and neuropathic response. These knockout mice revealed their respective mechanism of action, as their cells did not release the mature pro-inflammatory cytokine IL-1beta from cells, despite IL-1 beta expression remaining constant. Contrary to gene knockout models, are gene knock-ins models. In gene knock-ins, genes not originally in the mouse are inserted, and subsequent disease protein is synthesized. These transgenic animals usually have a different phenotype to a knockout and may mimic more closely what happens during disease and treatment.

Another in vivo technique used for target identification is antisense oligonucleotide-based models. Antisense oligonucleotides mimic RNA and complement the target mRNA molecule [12]. Bound antisense oligonucleotide prevents ribosomal translation of mRNA to protein. Honore and colleagues created an antisense oligonucleotide that inhibited translation of the rat P2X3 receptor [13]. When rat models were dosed with P2X3 antisense, they displayed anti-hyperalgesic activity. Once administration of the antisense oligonucleonucleotides was discontinued, receptor function and algesic responses returned. Unlike transgenic model, the antisense oligonucleotide effect is reversible [14].

While there are many viable target validation methods, two modern technologies can enable tissue specific validation: clustered regularly interspaced short palindromic repeats (CRISPR), and CRISPR-related techniques, and organs on a chip.

The CRISPR-Cas9 and related approaches provide multiple advancements compared to the transgenic model; these include the ability to overcome embryonic lethality and avoid resistance mechanisms. In brief, CRISPR-Cas9 works by distributing the Cas9 nuclease into the cell. Synthetic guide RNA then guides the nuclease to the desired cut location, facilitating the addition or removal of genes in vivo [15]. An example of CRISPR-Cas 9 target validation is with the elucidation of the mechanism of action behind tumour suppressor, p53, reactivating compounds. Employing CRISPR-Cas9-based target validation in lung and colorectal cancer displayed that the anti-proliferate activity of nutlin is dependent on functional p53. However, using traditional models, the mechanism and therapeutic response to p53-reactivating compounds is lost via compound-specific resistance mechanisms [16].

Another emerging technology that will facilitate improved target validation is organs-on-chips. These are multi-channel 3-D microfluidic cell culture chips that mimic the functionality and physiology of entire organs. This technology yields the potential to quickly assess the efficacy and human response to target mediation. Song and colleagues used a vasculature system chip model to assess the relationship between vascular endothelium and the metastatic behavior of circulating tumour cells. This study suggested that the inhibition of CXCL12-CXCR4 binding on endothelial cells may be a valid target in the prevention of metastasis [17]. Importantly, organs-on-chips technologies may provide novel insights to target identification and validation studies.

Overall, there are many means to validate targets; all strategies have a common aim: to evaluate the target's cellular function prior to full investment into the target, and drug candidate screening.

2.2 Drug discovery and development pipeline: lead discovery

Once the biological targets have been identified and validated, the next fundamental step is the lead discovery phase. This comprises of three components, in order: hit identification, hit-to-lead phase, and lead optimization.

2.2.1 Drug discovery and development pipeline: hit identification

It is during this phase that drug compound screening assays are developed, and subsequent 'hit' compounds derived. The term 'hit' compound is used in a range of terminologies; in this overview we refer to it as a compound that obtains the desired screening effect, which has been validated upon retesting. Various screening approaches exist to identify hit molecules. In this overview, we will describe the most common screening strategies: high throughput screening, Focused based screening, and fragment screening.

High throughput screening utilizes an entire compound library and assesses the activity of each compound on the biological target. This typically involves large semi-automated cell-based assays. A candidate hit compound typically requires further assays to confirm its mechanism of action [18].

Focused based screening, also termed knowledge-based screening, selects compounds from a library based on existing information about the target, stemming from literature or patents, which suggest compounds likely to yield the desired target activity [19].

Fragment screening uses small-molecular weight compound libraries and screens these compounds at high concentrations. Small fragments that bind to the target are often scaled with chemical alterations to increase their binding affinity [20].

2.2.2 Hit-to-lead phase and lead optimization

The aim of this intermediate phase is to develop a compound(s) with enhanced properties, with pharmacokinetics suitable for one or many different in vivo models. This step regularly involves a series of structure-active-relationship (SAR) investigations for each hit compound, in an attempt to measure the activity and selectivity of each compound.

The goal of the final lead discovery phase is to obtain compounds with optimal structural, metabolic, and pharmacokinetic properties. This often involves further applications of various in vitro and in vivo screens.

2.3 Drug discovery and development pipeline: preclinical

Once a lead candidate is identified, further elucidation of its structure, metabolic, and pharmacokinetic properties may be required. The typical preclinical development stage is comprised of various components, typically used with animal models: (1) The first preclinical experiments revolve around dose design; a safe dose must be identified with estimated human measurements. (2) Second, the pharmacodynamics of a compound is required; the mechanism of action that causes the clinical response, with respect to doses, must be determined. (3) Third, pharmacokinetics properties of the drug candidate are required. This includes absorption, distribution, metabolism, excretion, and potential drug-drug interactions. The aim of preclinical studies is to obtain enough information to determine a safe dose for the first human study. On average, one in 5000 preclinical development candidate drugs make it through preclinical development and become regulatory approved [21].

2.4 Drug discovery and development pipeline: clinical development

The clinical development/clinical trial stage comprised of three main stages and one post-market surveillance stage.

The phase 1 clinical studies are carried out in a small number of healthy volunteers. The aim of this stage is to distinguish a therapy's metabolic and pharmacological effects, as well as the side effect response to varying dosages. The main aim of phase 1 is to determine a therapy's safety profile.

Stemming from the data collected during phase 1, phase 2 studies also termed 'therapeutic exploratory' trials involve investigations on several diseased individuals. This phase aims to further determine the effectiveness of the drug with respect to disease or condition. Side effects and risks are further distinguished. Phase 2 studies are controlled, usually conducted on a few hundred patients.

The phase 3 studies are a much larger drug assessment of the drug's efficacy, safety, and evaluate the overall benefit-risk relationship of the drug. This phase may also yield enough data to estimate the results of a general population, as they include several hundred to several thousand people.

Once the drug is approved, there is a fourth phase, known as post-marketing surveillance. These are observational studies, whereby the goal is to define and ensure the safety profile of the drug on a larger population scale.

2.5 Drug discovery and development pipeline: challenges and overview

There are three main reasons why drugs fail: the first is that they simply do not work, second is that they are unsafe for clinical use, and the third reason of drug failure is due to poor clinical trial structure. The cost of a candidate soars the further it gets in the drug development pipeline.

The primary source of trial failure is a drug's lack of efficacy. Hwang and colleagues investigated 640 phase 3 trials, of which 54% failed. Over 50% of these failures were due to a lack of efficacy [22]. There are a variety of reasons why a drug may enter phase 3 trials and yet lack efficacy. This may also include the propagation of error due to flawed target validation, a poor study design, or simply having an insufficient number of patient trials resulting in weak statistical power and an inability to reject the null hypothesis.

The infamous, poly ADP ribose polymerase (PARP) inhibitor, Olaparib failed its first trial for ovarian cancer due to a lack of trial structure. In the initial trial, in individuals with the BRCA mutation and platinum-sensitive recurrent ovarian cancer, Olaparib delayed the time to recurrence to 11.2 months from 4.3 months. However, the median time to death was 34.9 months in the treatment group and 31.9 months in the control group ($p = 0.19$) [23]. In 2014, Olaparib was approved by the FDA for women with recurrent ovarian cancer who have the germline *BRCA* mutation and had previously received three or more lines of chemotherapy. This approval was based on a study by Kaufman and colleagues [24], which displayed a response rate > 30% with Olaparib monotherapy in patients who had previously received three or more lines of chemotherapy.

Clinical trials also fail with respect to safety. In Hwang and colleagues' study, out of the initial 640 compounds, 17% of them failed due to safety [22]. Drug safety is a key factor in every stage of the candidate drug development; however, challenges may only present at larger populations [25]. One reason for failure due to safety is due to ill reporting of safety concerns. Generally, a patient's safety concerns may not align with that of the administering physician. It is logical to assume people will be more likely to report an adverse event that is of concern to them. It is important that at each step within the drug development pipeline safety is a primary consideration. The cost of determining a safety issue propagates with progression through each drug development stage.

One of the most impactful drug candidate failures was with sulphanilamide. This drug was popular in the 1930s and sold in both a bolus and elixir form. However, important safety tests had not been conducted for the elixir form,

although at the time this testing was not required. Unfortunately, after being treated with the elixir form, over 100 people died due to diethylene glycol poisoning [26]. This led to the implementation of two important acts: The Food, Drug and Cosmetic Act and Drugs and Cosmetics Act.

3. The potential of AI

AI has been utilized in drug discovery since the early 1960s. However, in 2016 many large pharmaceutical companies started investing in AI by partnering with AI startups or academic groups or initiating their own internal AI R&D programs. This has resulted in an enormous number of new publications within the field that cover the entire drug discovery and development pipeline. This has included the implementation of deep learning models to predict the properties of small molecules from transcriptomics data [27] to the identification of novel drug targets [28]. AI has integrated into almost every area of drug discovery and development.

The primary aim of drug discovery and development combined with AI remains to facilitate the development of the best drugs and bring them to the clinic to fulfill unmet medical needs.

AI and machine learning has a lot of potential. For those new to the field, AI limitations seem endless, regardless of the input information. AI has a range of applications. It can be successful at creating an image of a cat from a model trained on images of cats or can enable a car to drive automatically without making a single mistake, or a drug that can be designed to treat a disease safely and efficaciously. However, AI will not succeed with every challenge; it is simply a tool that may drive new technologies, and enhanced understanding. In drug discovery and development, AI is not one entity that can design a drug from start to finish, but many different AIs which enhance our understanding throughout the drug discovery and development process.

3.1 Fundamental AI concepts

While many computational approaches can fit the broad definition of AI, two fields are currently popular: machine learning and its subfield deep learning. In layman's terms, the key difference with deep learning is that it uses multiple layers, each employing different calculations on the initial data. In order to understand their capacities, a few fundamental concepts must be understood.

Broadly, there are two different types of machine learning to understand. Supervised learning is when a model is trained using labeled data sets to predict a certain outcome. An example of this is the quantitative structure–activity relationship (QSAR) approach. This is used to predict a chemical's property, such as solubility and bioactivity [29]. The other approach is unsupervised learning, as the name suggests, it does not depend upon training with labeled data to find relationships with data. Examples include the use of hierarchical clustering, algorithms and principal components analysis to analyze and group large molecular libraries into smaller sub-groups of similar compounds.

With supervised machine learning, there are two types: classification and regression. Classification models are used when the problem is categorical, as in the predicted output is a limited set of values. Regression models are used when the problem involves predicting a numeric value within a range.

There are a variety of different types of machine learning models, such as random forests, autoencoders, and convolutional neural networks. Each of the subsequent chapters will describe specific models as required.

3.2 Examples of AI implementations in drug discovery and development

A vast number of AI and drug discovery papers are published every day, covering various aspects of the entire drug discovery and development pipeline. Drug discovery and development-based AI technologies range from the identification and validation of drug targets, drug repurposing, identification of new compounds, and improving the R&D efficiency. There are a number of potential contributions AI can make to reduce inefficiencies in the conventional drug development and discovery pipeline.

Target identification and validation have been enhanced by AI. This is made possible by genomics, with biochemical and histopathological information. The IBM Watson identified five novel RNA-binding proteins as potential targets linked to the pathogenesis of amyotrophic lateral sclerosis, which currently has no known cure [30].

One huge opportunity for AI in drug discovery is with drug repurposing. As an example, Donner and colleagues [31] used a transcriptomics data set and derived a new measurement of compound functionality, based on gene expression. This measurement allowed the identification of compounds that shared biological targets, despite being structurally different, revealing previously unknown functional associations between compounds.

An AI platform that can predict a candidate's mechanism of action and in vivo safety would cut wasted costs dramatically. There are several examples of companies with this goal. This includes DeeoTox and ProCTOR, both of which aim to predict the toxicity of new compounds [32, 33]. The performance of these AI platforms is expected to increase as larger robust data sets on the toxicity of compounds are made available.

As of 2019, one important study was the discovery of a drug within 21 days. Deep learning enabled the identification of potent DDR1 kinase inhibitors within 21 days. Out of the four compounds discovered, one lead candidate has displayed ideal pharmacokinetics in mice [34].

Overall, it is clear AI may yield increases in drug discovery efficiency through various strategies.

3.3 Current challenges in AI

AI has shown promise in drug discovery and development. However, it is not without its challenges. There are many challenges faced by AI in medical research such as lack of data, lack of interoperability, and the curse of dimensionality.

The lack of data is a recurring problem throughout every industry wanting to implement AI. The minimum number of samples in a traditional biological study is five, for it to be valid. However, most machine learning algorithms must be trained on hundreds, or thousands, of data points/samples, in order to perform well. Furthermore, obtaining labeled data can be a challenge, as this often requires some form of manual input. Fortunately, large databases, such as The Cancer Genome Atlas program (TCGA), are aggregating and open-sourcing vast amounts of robust data from multiple institutions. However, on some occasions, large databases that include the requested data may not exist. One such strategy to combat this is data augmentation. Data augmentation is the process of creating artificial data from real data. There are a variety of data augmentation approaches; ultimately they increase the data available for training models, without collecting new data.

Another challenge faced by machine learning is the lack of interpretability. The term 'black box model' is often used when it is difficult to explain how a model makes certain predictions and performs. This is more likely an occurrence with deep learning, as each layer adds complexity to the model explaining each layer's outputs

can become exponentially complex and the number of layers increases. However, a variety of tools are being developed in order to elucidate further explainability such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive Explanations). LIME adopts a local linear approximation of the model's behavior, whereas SHAP employs a game theory-based approach to explain the model output. Both LIME and SHAP, and other similar strategies, are projected to become common practice in machine learning and are going to be necessary to get more AI technologies to the clinic [35].

A recurring issue with artificial intelligence in medical data is known as the curse of dimensionality. This is when the data sets used have a small number of samples and many features. This is a common occurrence in medical omics data sets, as they typically yield thousands of features and less than 100 samples; thus the available data become sparse. This problem may be addressed with a variety of dimensionality reduction techniques.

Overall, there are a series of challenges that will need to be addressed for AI to reach its optimal capacity. In this passage, we have only described a few challenges. However, they are being addressed with advancements in complementary data science approaches and tools, such as the creation of large data repositories, tools to increase explainability, and the creation of feature reduction techniques.

4. Concluding remarks

Taking a drug from idea to the clinic is a long diverse process, costing over 2.6 billion dollars, and take over a decade to develop a cancer therapeutic. This is primarily due to high numbers of candidate drugs failing at late drug development stages. Advancements in AI are continually displaying the possibility of rapid low-cost drug discovery and development. As we make our way through the 2020s, it is evident the drug discovery and development will be permanently shaped by AI.

Acknowledgements

The author would like to thank N.B.N and I.H.

Conflict of interest

The author declares no conflict of interest.

Acronyms and abbreviations

AI	artificial intelligence
CRISPR	clustered regularly interspaced short palindromic repeats
ML	machine learning
NAT1	*N*-acetyltransferase 1
NAT2	*N*-acetyltransferase 2
PARP	poly ADP ribose polymerase
TCGA	The Cancer Genome Atlas

IntechOpen

IntechOpen

Author details

Kristofer Linton-Reid
Imperial College London, London, UK

*Address all correspondence to: kl2418@ic.ac.uk

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. BY–NC

References

[1] Avorn J. The $2.6 billion pill—Methodologic and policy considerations. New England Journal of Medicine. 2015;**372**:1877-1879. DOI: 10.1056/NEJMp1500848

[2] Seyhan AA. Lost in translation: The valley of death across preclinical and clinical divide—Identification of problems and overcoming obstacles. Translational Medicine Communications. 2019;**4**:1-19. DOI: 10.1186/s41231-019-0050-7

[3] Tripathy RK, Mahanta S, Paul S. Artificial intelligence-based classification of breast cancer using cellular images. RSC Advances. 2014;**4**:9349-9355. DOI: 10.1039/c3ra47489e

[4] Zhou Z, Li X, Zare RN. Optimizing chemical reactions with deep reinforcement learning. ACS Central Science. 2017;**3**:1337-1344. DOI: 10.1021/acscentsci.7b00492

[5] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Science Advances. 2018;**4**:eaap7885. DOI: 10.1126/sciadv.aap7885

[6] Hofmarcher M, Rumetshofer E, Clevert DA, Hochreiter S, Klambauer G. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. Journal of Chemical Information and Modeling. 2019;**59**:1163-1171. DOI: 10.1021/acs.jcim.8b00670

[7] Klambauer G, Hochreiter S, Rarey M. Machine learning in drug discovery. Journal of Chemical Information and Modeling. 2019;**59**:945-946. DOI: 10.1021/acs.jcim.9b00136

[8] Yin Z, Ai H, Zhang L, Ren G, Wang Y, Zhao Q, et al. Predicting the cytotoxicity of chemicals using ensemble learning methods and molecular fingerprints. Journal of Applied Toxicology. 2019;**39**:1366-1377. DOI: 10.1002/jat.3785

[9] Hein DW. Molecular genetics and function of NAT1 and NAT2: Role in aromatic amine metabolism and carcinogenesis. Mutation Research. 2002;**506-507**:65-77. DOI: 10.1016/s0027-5107(02)00153-7

[10] Golka K, Prior V, Blaszkewicz M, Bolt HM. The enhanced bladder cancer susceptibility of NAT2 slow acetylators towards aromatic amines: A review considering ethnic differences. Toxicology Letters. 2002;**128**:229-241. DOI: 10.1016/s0378-4274(01)00544-6

[11] Kurosawa G, Akahori Y, Morita M, Sumitomo M, Sato N, Muramatsu C, et al. Comprehensive screening for antigens overexpressed on carcinomas via isolation of human mAbs that may be therapeutic. Proceedings of the National Academy of Sciences of the United States of America. 2008;**105**:7287-7292. DOI: 10.1073/pnas.0712202105

[12] Taylor MF, Wiederholt K, Sverdrup F. Antisense oligonucleotides: A systematic high-throughput approach to target validation and gene function determination. Drug Discovery Today. 1999;**4**:562-567. DOI: 10.1016/S1359-6446(99)01392-6

[13] Honore P, Kage K, Mikusa J, Watt AT, Johnston JF, Wyatt JR, et al. Analgesic profile of intrathecal P2X3 antisense oligonucleotide treatment in chronic inflammatory and neuropathic pain states in rats. Pain. 2002;**99**:11-19. DOI: 10.1016/S0304-3959(02)00032-5

[14] Miller CM, Harris EN. Antisense oligonucleotides: Treatment strategies and cellular internalization. RNA & Disease. 2016;**3**(4):e1393. DOI: 10.14800/rd.1393

[15] Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nature Biotechnology. 2015;**33**:985-989. DOI: 10.1038/nbt.3290

[16] Wanzel M, Vischedyk JB, Gittler MP, Gremke N, Seiz JR, Hefter M, et al. CRISPR-Cas9-based target validation for p53-reactivating model compounds. Nature Chemical Biology. 2016;**12**:22-28. DOI: 10.1038/nchembio.1965

[17] Song JW, Cavnar SP, Walker AC, Luker KE, Gupta M, Tung Y-C, et al. Microfluidic endothelium for studying the intravascular adhesion of metastatic breast cancer cells. PLoS ONE. 2009;**4**:e5756. DOI: 10.1371/journal.pone.0005756

[18] Entzeroth M, Flotow H, Condron P. Overview of high-throughput screening. Current Protocols in Pharmacology. 2009. Chapter 9: Unit 9.4. DOI: 10.1002/0471141755.ph0904s44

[19] Boppana K, Dubey PK, Jagarlapudi SARP, Vadivelan S, Rambabu G. Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. European Journal of Medicinal Chemistry. 2009;**44**:3584-3590. DOI: 10.1016/j.ejmech.2009.02.031

[20] Price AJ, Howard S, Cons BD. Fragment-based drug discovery and its application to challenging drug targets. Essays in Biochemistry. 2017;**61**:475-484. DOI: 10.1042/EBC20170029

[21] Umscheid CA, Margolis DJ, Grossman CE. Key concepts of clinical trials: A narrative review. Postgraduate Medicine. 2011;**123**:194-204. DOI: 10.3810/pgm.2011.09.2475

[22] Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. JAMA Internal Medicine. 2016;**176**:1826-1833. DOI: 10.1001/jamainternmed.2016.6008

[23] Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: A preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. The Lancet Oncology. 2014;**15**:852-861. DOI: 10.1016/S1470-2045(14)70228-1

[24] Kaufman B, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, Balmaña J, et al. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. Journal of Clinical Oncology. 2015;**33**:244-250. DOI: 10.1200/JCO.2014.56.2728

[25] Crowther M. Phase 4 research: What happens when the rubber meets the road? Hematology/The Education Program of the American Society of Hematology American Society of Hematology Education Program. 2013;**2013**:15-18. DOI: 10.1182/asheducation-2013.1.15

[26] Paine MF. Therapeutic disasters that hastened safety testing of new drugs. Clinical Pharmacology and Therapeutics. 2017;**101**:430-434. DOI: 10.1002/cpt.613

[27] Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Molecular Pharmaceutics. 2016;**13**:2524-2530. DOI: 10.1021/acs.molpharmaceut.6b00248

[28] Li Q, Lai L. Prediction of potential drug targets based on simple sequence properties. BMC

Bioinformatics. 2007;**8**:1-11. DOI: 10.1186/1471-2105-8-353

[29] Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: Where have you been? Where are you going to? Journal of Medicinal Chemistry. 2014;**57**:4977-5010. DOI: 10.1021/jm4004285

[30] Bakkar N, Kovalik T, Lorenzini I, Spangler S, Lacoste A, Sponaugle K, et al. Artificial intelligence in neurodegenerative disease research: Use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. Acta Neuropathologica. 2018;**135**:227-247. DOI: 10.1007/s00401-017-1785-8

[31] Donner Y, Kazmierczak S, Fortney K. Drug repurposing using deep Embeddings of gene expression profiles. Molecular Pharmaceutics. 2018;**15**:4314-4325. DOI: 10.1021/acs.molpharmaceut.8b00284

[32] Gayvert KM, Madhukar NS, Elemento O. A data-driven approach to predicting successes and failures of clinical trials. Cell Chemical Biology. 2016;**23**:1294-1301. DOI: 10.1016/j.chembiol.2016.07.023

[33] Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: Toxicity prediction using deep learning. Frontiers in Environmental Science. 2016;**3**:80. DOI: 10.3389/fenvs.2015.00080

[34] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology. 2019;**37**:1038-1040. DOI: 10.1038/s41587-019-0224-x

[35] Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 13-17, August-2016, Association for Computing Machinery. 2016. pp. 1135-1144. DOI: 10.1145/2939672.2939778