

Type of the Paper (Article, Review, Communication, etc.)

Concept of Artificial Intelligence in discovering and re-purposing of drugs

Sushant Bajpai^{b†}, Nehil Shreyash^{a†}, Muskan Sonker^a, Vishwas Gupta^b, Saurabh Kr Tiwary^{a*}, Susham Biswas^{a*}

¹ Affiliation b; Department of Petroleum Technology, Rajiv Gandhi Institute of Petroleum Technology

² Affiliation a; Department of Chemical Technology, Rajiv Gandhi Institute of Petroleum Technology

* Correspondence: susham@rgipt.ac.in

† Equal Contribution

Abstract: Artificial Intelligence (AI) is a platform lending immense assistance in discovering and developing drugs and thus, various such approaches have been developed with the intent of simplifying and improving biomedical operations such as drug repurposing and drug discovery. In the past decade, AI-based investigation of nanomedicines, as well as non-nanomedicines has reached the clinical level. In semblance with the traditional methods of therapy, nanomedicine therapy is employed at limited doses. The study of a variety of drugs resulted in the conclusion that the effect of each drug is variable for every patient and, evaluating that perfect drug combination manually is a time-consuming as well as an inefficient treatment method. Therefore, the use of AI simplifies and reduces the time consumption in determining the perfect customized drug combination for nano-therapy. The area with the most potential for meeting this reality is to optimize the drug and dosage parameters. It is a universally known fact that cancer is dangerous and unique because of the exacting challenges it poses during treatment and, to achieve a better treatment, the therapeutic effect on each patient must be delineated even if the volume of data generated is massive. The article aims at analyzing the AI technologies that help yield results much quicker, make the analyses simple, and efficient.

Keywords: Artificial Intelligence, Drug Delivery, Cancer, Nanomedicines, Therapeutic

1. Introduction

The research and development (R&D) pertaining to small-molecule drugs are presently confronting many challenges like increased cycle time, unsuccessful clinical trials, and some economic issues. Even with ample sponsoring, the drug (small-molecule) R&D is yet to yield the necessary productivity and efficacy. Elongated approval procedures for incorporation of novel chemicals and techniques, saturation of market along with hesitation to invest in developing and developed markets are issues that need addressal. In this review, we shall be discussing challenges faced from the discovery of drugs to initial clinical trials. In recent times, amount of information available to scientists is incomparable, sufficient to outpace abilities of many conventional practices. A prominent solution to the aforementioned challenges may be achieved with Machine Learning and optimum utilization of artificial intelligence (AI) which might not be the case as of now. Thus, the use of machine intelligence is offering opportunities for combating challenges regarding small-molecule drugs. There has been an evolution in fundamentals and algorithms of Machine learning tactics which have a very narrow area of processing and may be considered as "weak" AI. Consequently, we shall be using the term AI for these Machine learning approaches because to date no such AI has been developed for wide range processing and may be categorized as "strong" (general). Hence in this review, we shall analyse variously available and proved useful techniques from this sub-domain as well as

some promising technologies for future exploration in the drug delivery domain.

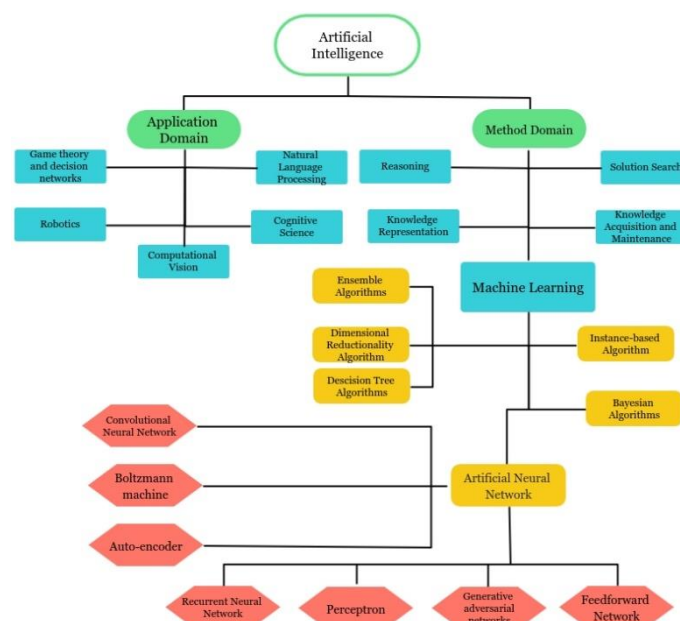


Figure 1: Artificial Intelligence Divisions

In recent times, the spike in AI is attributed to exploitation of stored data or Big data, associated advancement in computer power (Google's tensor processing units (TPUs), Graphics processing units (GPUs)), and ML algorithms optimization. Training of nontrivial network elements combinations, considering huge data size, in optimum duration, resulting in elevated practicality of these models are now attainable. Thus, its potential for training deep hierarchical networks within plausible time frame portrays its capabilities for extensive application in areas like medical imaging, autonomous vehicle, and many more.

AI is not a new concept for drug discovery domain, particularly modelling in generalized structure-activity relationships. Hammett's formula building a connection between equilibrium constants and reaction rates for benzene and its derivatives is analogous to the idea of considering experimental values and proceeding for regression.

Since then, a lot of efforts have been made to predict the biological impact of various chemical compounds by employing AI technology. Pattern recognition approach, for example, focuses on scrutinizing and elucidating to discover a relation between chemical structures given the fact that often chemicals with same structures share the same in-vitro biological and physicochemical properties. Several neural networks implemented models and prototypes developed showed capabilities to solve such issues.(elaborate issues?)

More advanced methods have evolved with the development of deep learning around 2010. The capabilities shown by deep learning models to analyze and predict connections shared by observations and molecular representations makes it a significant frontrunner for more accurate and generalized predictions. Owing to their exceptional potential these methods won challenges like compound activity prediction (2012) and the NIH Tox21 challenge for the prediction of toxicity (2014). These achievements have channeled pharmaceutical companies to collaborate with leading AI for advanced results.

Problem definition	<ul style="list-style-type: none">• Discrimination (e.g. involves the division of molecules into categories of active and inactive) or• Generation (e.g. requires the creation of a molecule comprising of the properties desired)
Design of AI architecture	<ul style="list-style-type: none">• Making an appropriate choice of an AI algorithm• Making an appropriate choice of initial values for the hyper-parameters
Input data preparation	<ul style="list-style-type: none">• Quality and quantity of data (e.g. complexity and type, distribution in chemical space)• Data pre-processing• Proportions of training and test sets (sometimes, a validation set is included)• Molecular representation
Model training and evaluation	<ul style="list-style-type: none">• Training algorithms (e.g. stochastic gradient descent, momentum, Nesterov Momentum)• Optimization strategies (e.g. batch normalization, supervised pre-training)• Evaluation mechanisms (e.g. hold-out validation, cross-validation, bootstrapping)• Evaluation metrics (e.g. area under ROC curve, mean absolute error, R^2)• Hyper-parameter tuning algorithm (e.g. grid search, random search, smart hyper-parameter tuning)

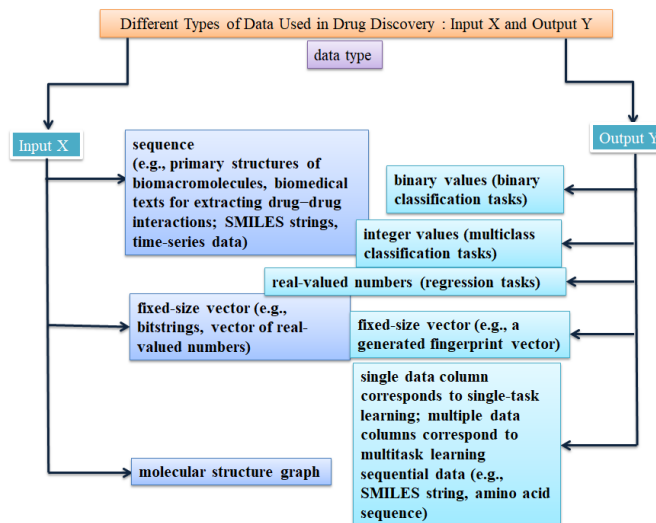
2. Role of Artificial Intelligence Modelling in Drug discovery

2.1 Construction of a model.

To be able to apply AI in drug discovery, the problem must be elucidated properly and with great accuracy². In the procedure for AI modelling, preparing the data sets, constructing the architecture for applying AI, etc are the most fundamental tasks (**Figure 2**). However, there must be clarity in understanding the problem even before the designing of the architecture because for a particular problem of interest a proper mechanism of machine learning is important to be selected. After there is sufficient knowledge about the problem under study, the time for the model design comes where a choice about the algorithm must be made. Also, in addition to this, analytically choosing initial values of hyperparameters is important. For carrying out discriminative tasks the use of algorithms like DBM (Deep Boltzmann Machine), VAE (variational autoencoder), or AAE (average absolute error) is mostly favored while Artificial Neural Networks (ANNs) are used to carry out the general tasks. Variation in Hyperparameters occurs with differing AI algorithms, once the designing of the model is complete, the preparation of model training and evaluation is carried out. With a target of minimizing the error incurred during prediction, the search for a parametric set is achieved in the training step.

2.2 Preparation of input data

The latest developments in AI methods for drug discovery aims at enhancing the performance of the model by being more focused on practical approaches in AI. Although, there are many more benefits of focusing on the training data first since all of the progress is dependent on the training data while preparation of data is a laborious and taxing task.



The basics of training data must be completely understood. A sufficient amount of training data must be available prior to³ dataset preparation and is a pivotal step that affects the efficacy of a model. The steps following this are

2.2.1. Types of data in medicinal chemistry

Figure 2 depicts different types of data that may be taken as input (X) or output (Y) to construct AI models for drug discovery and fixed-length input vector is one of the most frequently used data type for input data⁴⁻⁶. Two disadvantages of this representation are that these vectors are too huge and it is difficult to encode all the substructures without overlapping⁷. Therefore, they become those models which have many learnable parameters. The latter limitation occur because of problems incurred in the establishment of a bijection between input vectors and molecular structures. It is easy to generate input vectors from molecular structures but processing of the same is extremely challenging. The use of AI deep generative models and SMILES (simplified molecular-input line-entry system) strings as molecular representations is a method by which the above limitations can be lifted⁸.

2.3 Applications of Artificial Intelligence in drug discovery for cancer

Li et al conducted an experiment with models created to understand the sensitivity of drugs and using the available data from the BATTLE patients. When all this was combined with the publicly accessible data from all over the world, the model was able to find out the sensitivity to each drug. Iorio et al checked 990 cancer lines of the cell with 265 drugs and tried to figure out the action of a number of variable quantities including methylation, mutation effects, gene copy number, and gene expression could cause some visible difference in the response of the drugs. The molecular features that actually affect drug response were found through a number of models. Gene expression was found to be the high impact quantity in the prediction. In a similar vein, RNA interference data was studied by Tsherinak et al. A regression was created which was non-linear and based on conditional interferences classifier for classification of proteins into targets for different types of cancers, and this resulted in the identification of about 122 cancer targets that would work in any cancer and 462,266 and 355 were the number of specific targets for specific cancers of breast, pancreas, and ovaries.

Some attempts at using shared data have also been made, NCI DREAM is one such example. A number of teams participated and used their best possible model on the provided data set. Profiling data and cell lines were available which were RNA microarray, RNA sequencing, single nucleotide polymorphism (SNP) array, exome sequencing, reverse phase protein array, and DNA methylation status. A number of approaches were applied to these dimensions of profiles. The results obtained from this experiment are relevant even now for validating new developments.

2.4 Applications of Artificial Intelligence in nanoparticles-based Drug Discovery

Development of a drug with the help of conventional technologies is reliant upon the discovery of newer compounds based on their targets and analysis of the compounds for evaluating the efficiency of the drug. During drug trials, the compounds with no desired effect during treatment are rejected. Many a times, it has been observed that a drug is not considerably efficient when used individually but is extremely useful in combination at a proportion with some other drug. So, the large number of drugs available and a large number of combinations for proportional permutations produces challenge. With the help of conventional technologies for developing a therapy it is very difficult to personalise and formulate the perfect combination for every individual thus, the role of AI becomes important for nanomedicines and non-nanomedicines. In oncology, the researchers aim to achieve the perfect combination of therapies only through the experiments that are carried out initially and, in many cases, this method has led to the development of successful potions but often they fail when tested on patients. A research-based on Quadratic Phenotypic Optimization Platform (QPOP), an AI platform, was conducted wherein huge numbers of drugs were delineated with the aim to find the perfect therapeutic combination that can prove its potential against bortezomib-resistant multiple myeloma. The optimization carried out by the platform resulted in concluding that the most efficient combinations for bortezomib-resistant multiple myeloma are decitabine and mitomycin. Although, when these therapeutics were used individually, their action was not satisfactory but, when in perfect combinatorial proportion they were highly effective. In this research, AI was employed for finding the perfect combination even before the drug was out for clinical use. The ability of the QPOP AI platform was commendable as it was able to ensure that the therapeutic combinations had large efficiency with minimal side-effects. The samples from different patients were also investigated with the help of QPOP platforms and the response was found to be highly individualistic (figure 3). The advantage offered by AI was achievement of results without delineating the complex mechanisms.

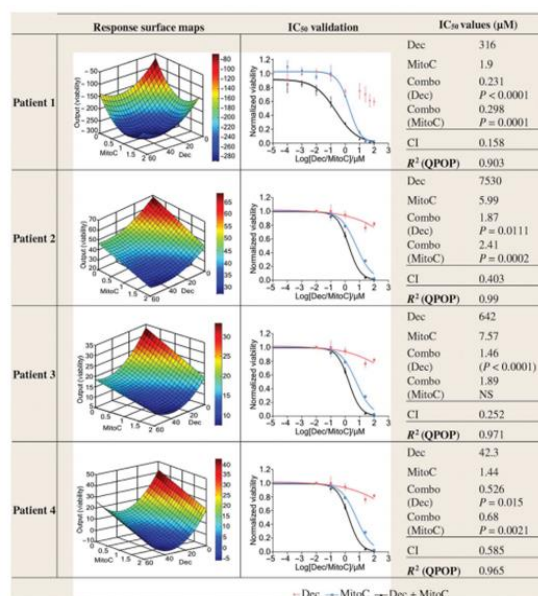


Figure 3: Artificial intelligence-driven optimization of novel combination therapies. Reprinted with permission from The American Association for the Advancement of Science²¹

3. Using the computational advancements in Deep Learning and Artificial intelligence for the repurposing of drugs

3.1 Employment of Artificial Intelligence technologies for repurposing the extant drugs

It is a commonplace fact that discovering a new drug requires huge manpower, capital, and time investment, and thus researchers have tried to put the existing drugs into use for a variety of conditions. Their main aim is to find the multiple applications of a particular drug and this is done by deeply assaying the existing drugs. This process is called Drug Repurposing. Opportunities for the repurposing of drugs exist mainly due to the fact that there are many target sites for a particular drug^{21,22} and distinct diseases may have some genetic similarity or follow a similar mechanism of action^{23,24}. By summing up the volume of data derived from different sources, the development of different AI-based methods has been done such as inference, indication and personalisation of novel drugs (PREDICT)³⁶, Simultaneously Localization and Mapping (SLAMS), Laplacian regularized least square method improved by a new kernel derived from interactive network between drug and protein (NetLapRLS)³⁷, and prediction of novel interaction between drug and target from heterogeneous constructed network (DTINet)³⁸. The AI technologies which are currently used can be majorly classified based on the learning style which is used in the construction of a model.

3.1.1. Repurposing the drugs under Supervised Learning

The AI-administered studies of drug repurposing done so far have majorly laid their focus on learning different algorithms for developing supervised predictors³⁹⁻⁴³. These supervised predictors are programmed on the basis of the already known relationships existing between the drugs, the drug targets, and the disease itself. The representation of these relationships is done with the help of different features like semblance in gene expression or the 2-D fingerprint etc. They are also represented by side-effects that happen due to a drug. The drugs that are repurposed with the aid of AI have an added advantage that they are tested on the basis of the replete amount of data available on even the smaller molecules, biological processes, etc. The huge amount of data is an advantageous fact while repurposing but it is also the root cause for problems like the integration of this data. Integrating a large amount of data is difficult as they originate from a variety of sources, making the data set heterogeneous. A technique was demonstrated by Napolitano et al³⁹. (**figure 4**) that offers a proper solution to the above-mentioned issue of data heterogeneity.

As can be seen in figure 5 that they made use of three kernels and used them to bring together the different sets of information.

Although the above-mentioned technique made us of kernels their use is not mandatory on all drug repurposing models that use AI. There are other techniques also available for integrating the

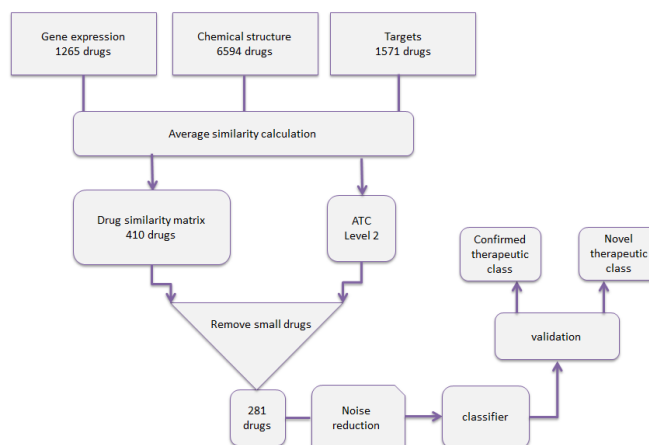


Figure 4- The figure is a flowchart that shows the analysis made by Napolitano et al. Three kernels were constructed in this for the purpose of pairwise computing the semblance between compound chemical structures, between the compounds on the basis of transcription profiles and between the target on the basis of a statistical analysis of a protein interaction network. The kernels were then used, to sum up, and form an overall

similarity matrix which is then analyzed with the help of an SVM classifier. The boxes in green are representations of data while the red boxes are those for the processes. Reproduced with permission from ref ⁴⁴(Creative Commons Attribution 2.0 International License: <http://creativecommons.org/licenses/by/2.0/>) Copyright 2013.

data as well such as the use of an approach laying its dependence on networks. In this, the data is aggregated relationship like that of gene expression, etc is carried out. This technique is best demonstrated by DTINet wherein for the drugs, four similarity networks were developed laying its basis interactions and relationships between the drugs, between the drug and the disease, between the side effects of the drug and on the chemical structures (Figure 5). together with the help of a network diffusion process, and from the distribution of this diffusion, an attempt to draw a

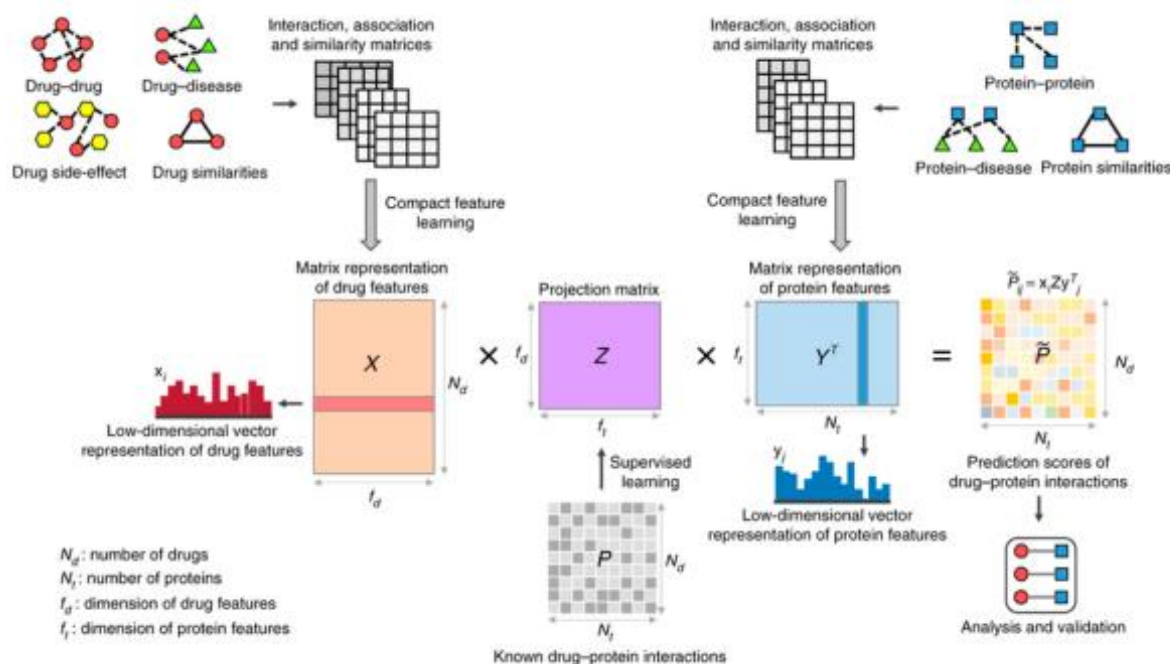


Figure 5- The figure is a representation of the DTINet pipeline in the form of a flowchart that provides an overview of the methods employed in the integration of heterogeneous information sources. Reproduced with permission from ref ³⁷(Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0/>) Copyright 2017.

3.2. Applying Deep Learning in drug repurposing and in predicting the pharmacological properties of drugs

Deep learning is a more refined form of machine learning⁵² that thrives on the capabilities of the newer, more intricate systems with humongous raw processing power and has displayed results with magnitude larger than all the research in the 2000s. It has improved performances by a little over 30% compared to a nominal 1-2% increase in the

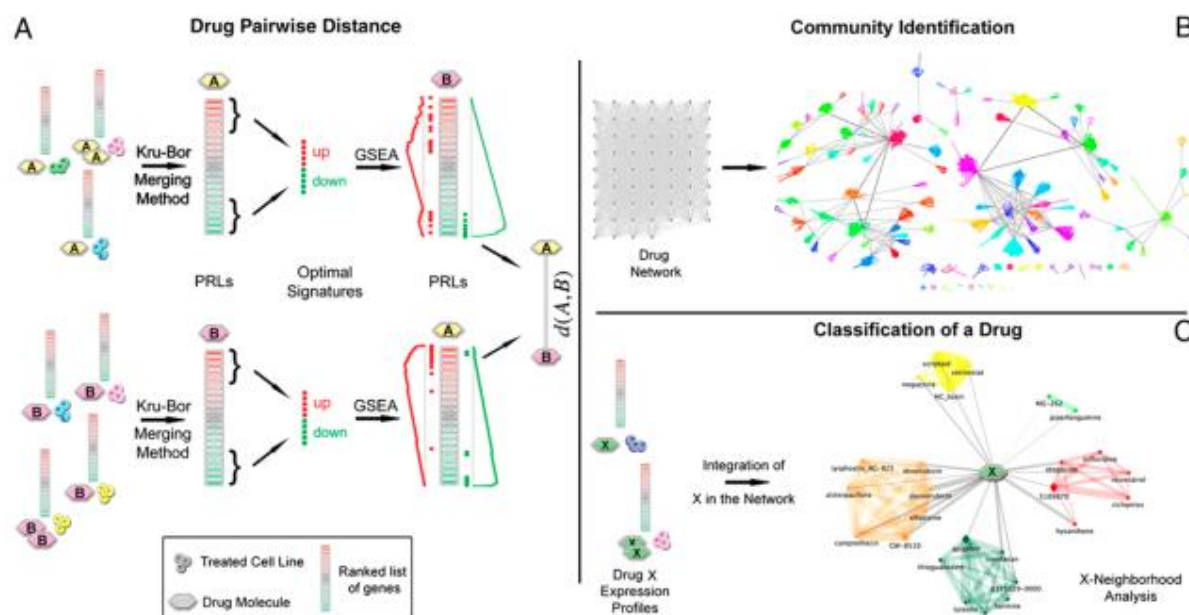


Figure 6- The figure depicts the methodology of the MANTRA approach. (A) With the help of rank-merging procedure and gene eset785 enrichment analysis, the determination of the distance between the pair of drugs was done. (B) The generation of a drug network was done wherein, each and every drug was considered to be a node having weighted edges where the pair of drugs were connected below a particular threshold value of distance. The identification of Network communities are carried out. (C) The drug X is that which has not been described and the evaluation of distance, $d(X, Y)$, was done for a drug Y. The addition of X in the network and this is described in (B). Reproduced with permission from ref ¹⁴⁶. Copyright 2010 National Academy of Sciences

whole of the decade preceding this advancement. Machine learning requires vast amounts of data and with the advent of those techniques, machine learning has even become capable of producing new data for the other models, paving way for a self-reliant learning method. The models have been able to extrapolate the three-dimensional structure of proteins using omics with the data derived before and after treatment about the genes and their expression. This, when combined with the available transcriptional profile of these drugs can produce accurate predictions of their transcriptional performance. One such study was performed by Alper et al⁵³ using 12 categories of therapeutic use. The samples used were from the LINCS (Library of integrated network-based cellular signatures) project including the A549, MCF-7, and PC-3. The training dataset was the pathway activation scores and gene expression level data.

The dimensionality in the data causes unexpected complexities; so, to counter that, pathway activation scoring was used⁵⁴. Certain genes were classified as the landmarks and became the new features. Through the pathway analyzer⁵⁵⁻⁶⁰, we might be able to figure out which pathways to preserve and use for further proceedings, on the whole, this reduces the dimensionality of the model. The landmark genes were decided by the data from the LINCS project, these particular landmarks hold more data than some of the other genes combined.

A great result that could be achieved from the data science capabilities is the correct reclassification of drugs. We are not fully aware of the interaction between the target and drug; thus, it is possible that some of the classifications might be wrong. With the validation from data analysis, these can be corrected. Otenzepad^{61,62} had been misclassified as a drug for the central nervous systems in the model, and it was only used in cardiac arrhythmia, but it has an important role in brain function, pinacidil was another example⁶³. After this, it was realized that this misclassification can be useful and allow for the development of the model as a drug repurposing and repositioning tool. The misclassification hence could be an indication for a new use of the drug.

3.3 Applications of AI in repurposing the drugs for Cancer Therapy

Drug discovery, repurposing, and repositioning still remains the crux of the use of

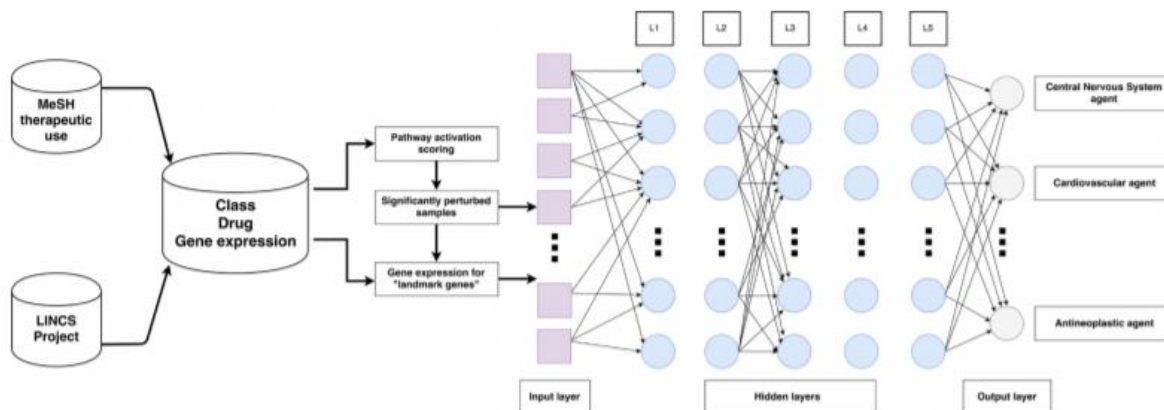


Figure 7: Use of artificial intelligence to identify the best-suited target points for drugs. In this particular case the data was taken from LINCS project and utilized with 12 therapeutic usage categories according to MeSH. Landmark genes were separately integrated to the DNN model. The total input layers equaled 977 and 271 neural nodes. Reprinted(adapted) with permission from ACS from reference⁵³ copyright(2016)

artificial intelligence. For the approval of a drug with conventional methods, a combination of the approved drug and investigative medicine was administered at the same time for attacking multiple targets. After this, the doses are altered and the effects observed of the best combination, the only limitation was the fact that if a drug went off-target, it won't be apparent. But, this kind of testing is simply the application of raw force and can actually be harmful in unprecedented ways, major downfall was its repetitiveness on different types of cancers, or diseases, in general. But with the advent of AI, the number of tests can be substantially reduced because of the superior analytic capabilities. The main objective of the multiple drug treatment was to find the perfect harmony between the drugs so that they can increase their efficiency. But with the knowledge from Data Science, it has been proven that even without a synergetic combination this goal can be reached. It has been shown in studies that if a number of drugs were to act independently with efficiency, it might actually give a better treatment outcome than a harmonious action of the drugs. With the help of this idea and the barrier of synergy out of the way, a newer class of drugs can be tested with the ML models. One of the major advantages is the freedom from needing to test every time the possible effect of drugs over one another. This tedious process is eliminated by the machine. It tests the combination of drugs and yields the best possible combination. With this method of quicker identification of combinations, some unpredictable combinations may be produced that is better than even the conventional set of problems. Every little margin that was considered too small for consideration can now be changed and the efficiency in the particular condition checked. Dosing can also be customized through a computer program to find the optimum dose of the drug.

The above is an example of repurposing Non-oncology drugs for cancer therapy. It is based on the progress made in the direction to fine-tune the drugs to individual molecular profiles rather than a generalized "One drug fits all" approach⁶⁴⁻⁶⁶. However, such applications require detailed data on genes and proteins derived from a systematic analysis of drug discovery for cancer therapy⁶⁷⁻⁶⁹. In the early turn of this century, cancer researchers Douglas Hanahan and Robert Weinberg presented the idea of "The Hallmarks

of Cancer". The authors believe that below the complexity of a diverse list of Cancer types lies a small number, or rather six underlying principles that govern the transformation of

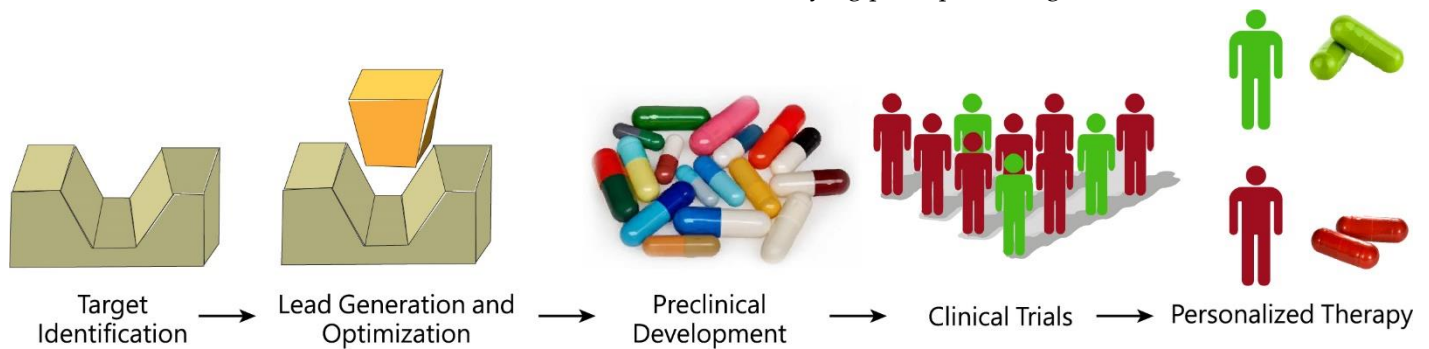


Figure 8: The aspects of cancer therapy to develop using artificial intelligence.

normal cells to cancer cells. The example presented above is aimed at targeting and disrupting these key hallmarks. It must be kept in mind that these key hallmarks of malignancy are not controlled by a single signal pathway which makes multi-hallmark targeting drugs advantageous as they are present in the vicinity to target multiple signaling pathways along with avoiding the possibility of adaptive resistance development⁷⁰⁻⁷².

These finely tuned drugs have the capability to dictate the complex signal interfaces and interactions that contribute to cancer cell generations. To name a few, CAFs cancer-associated fibroblasts, CTLs cytotoxic T lymphocytes, ECM extracellular matrix, MDSCs myeloid-derived suppressor cells, NK cells natural killer cells, Tregs regulatory T cells.

Specially targeted drugs with Anti-cancer activity accompanied

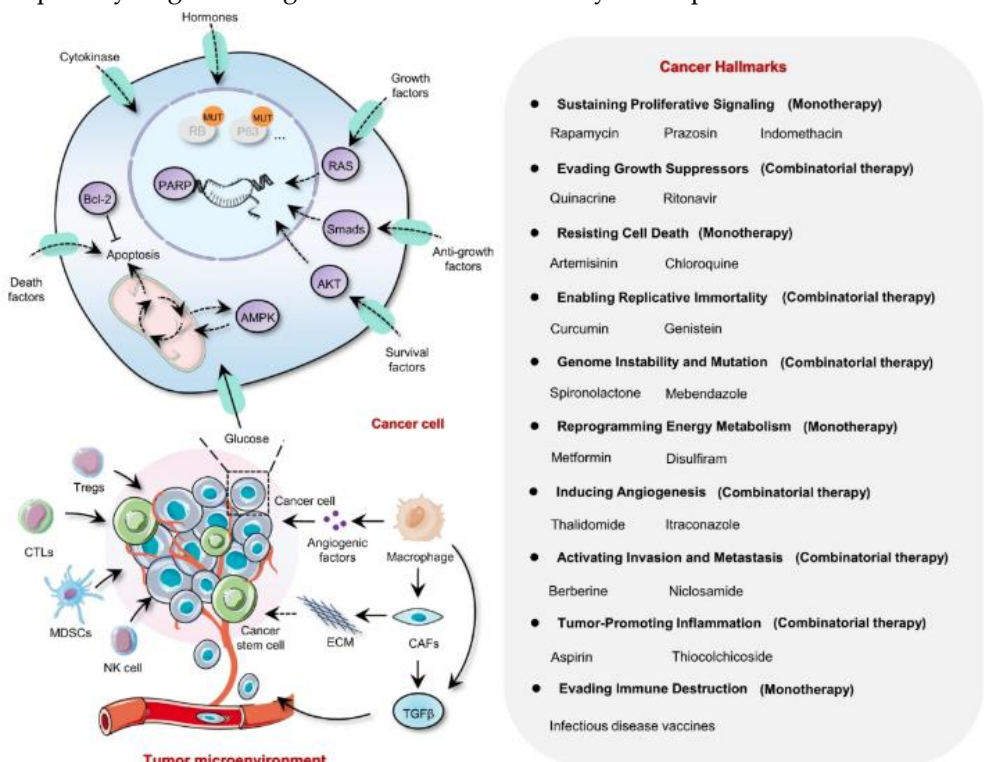


Figure 9 The possible drugs for which are suitable treatments for cancer can be identified with the help of drug repurposing and can be enabled with the help of recapitulative signalling networks. The complex signalling interactions that are major contributors to the hallmarks of cancer can be engineered and the complexities related to the neoplastic disease can also be rationalized. The drug candidates that have the capacity to fight against

cancer are mentioned. CAFs cancer-associated fibroblasts, CTLs cytotoxic T lymphocytes, ECM extracellular matrix, MDSCs myeloid-derived suppressor cells, NK cells natural killer cells, Tregs regulatory T cells Reprinted with permission from Springer Nature⁷³, the article is open access and the content reusable. Creative Commons International License 4.0 for the article is available from <http://creativecommons.org/licenses/by/4.0/>

by a very few off-target effects have played major roles in increased chances of survival of a cancer patient. But, often the parallel signaling⁷⁴ pathways have adapted to therapy pressure and in turn, returned to their function of aiding malignancy which ultimately leads to relapse by the transitory tumor therapeutic effect. All evidences lead to a revelation of

adaptive resistance caused by targeted drugs. Moreover, this major shortcoming pushed harder for the discovery of traditional non-oncology drugs having the ability to inhibit more than one hallmarks⁷⁵.

The first example depicted in Figure 10 depicts the signaling pathways mediated by aspirin. The aspirin orcheseters several signaling pathways amongst the hallmarks such as tumor-promoting inflammation, deregulating energy metabolism, angiogenesis, cancer metastasis, and immune evasion by regulating different target proteins like COXs, PP2A, RAS, IKK β , and histone.

Figure 11 depicts the signaling pathways mediated by Metformin mechanized by its direct and indirect impact on protein. Metformin acts on several signaling pathways including insulin/IGF1, NF- κ B, AMPK/mTOR/PI3K, Ras/Raf/Erk, Wnt, Notch, and TGF- β signaling, which in turn lead to cell proliferation, self-renewal, cancer metastasis, angiogenesis, and energy metabolism Research evidence has shown positive results for the role and impact of repurposed drugs for cancer treatment such as antifungal agents (ciclopirox olamine, ketoconazole, and itraconazole), antiparasitic drug (ivermectin), anthelmintic drug (pyrvinium), antibiotic agent (brefeldin A), and natural products (quercetin, 3'-epi-12 β -hydroxyfroside, and toxicarioside O). Figure 12 depicts the impact of a few such drugs namely itraconazole, ivermectin, ketoconazole, and brefeldin A. Brefeldin A, which is characterized as a macrolide antibiotic causes the death of an autophagic cell in both vitro and vivo in colorectal cancer. The ketoconazole is a traditional antifungal drug that induces PINK1/Parkin-mediated mitophagy by downregulating COX-2 causing the acceleration of apoptosis which plays a major role to inhibit the growth of HCC. The Ivermectin induces PAK1/Akt regulated cytosstatic autophagy which acts against breast cancer cells⁷⁶ Ivermectin has also displayed actions against the SARS-CoV-2 in vitro according to recent studies. Itraconazole has strong actions to counter brain cancer. A synopsis of the research carried out gives the scientific community further motivation to develop the strategy of repurposing non-oncology drugs for cancer treatment.

4. Artificial Intelligence technologies on other branches of drug delivery

4.1 Use of Data Analysis for Precision Medicine

In the present scenario, the fundamentals of finding out new drugs aren't simply based on the idea of mitigating the symptoms of the disease. They are more focused on finding out the root cause of the disease and attempt to completely eradicate or at least

inhibit its development in the host. One of the reasons of its achievability is due to the existence of mechanisms to monitor the subject through

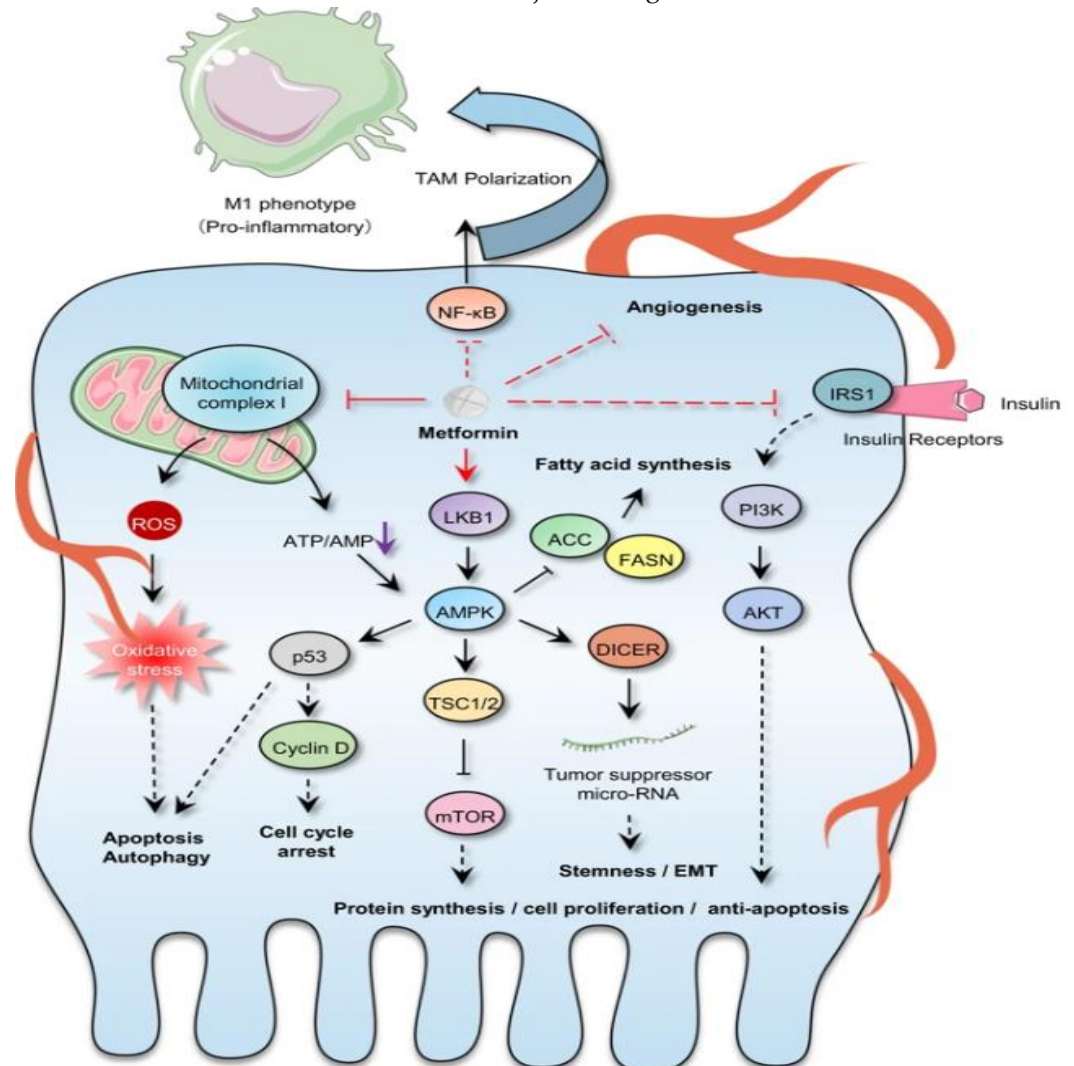


Figure 10: Signalling pathways mediated by aspirin. The various signalling pathways regulated aspirin are shown, along with acting on diverse hallmarks of cancer including tumor-promoting inflammation, deregulating energy metabolism, angiogenesis, cancer metastasis and immune evasion. Reprinted with permission from Springer Nature⁷³, the article is open access and the content reusable. Creative Commons International License 4.0 for the article is available from <http://creativecommons.org/licenses/by/4.0/>

devices that stay close to the body of the host all the time. They can figure out the behavioral and physiological data. This data, when combined with the available records and the biological data, yields a new class of customizable medicine for the individual. The targeting capability of the drugs also increases with the advent of these technologies as they provide us with the origins of the disease, thus producing precision medicine⁷⁷

In this era of the development of Artificial Intelligence, another challenging aspect has been, to a certain extent, made easier. The data banks described above are extremely large in their contents and also very heterogenous in definition and structuring. Also with the advancement of communications across the planet contributions are very quick leading to information overload. It would have been a humongous task to scan all the data and keep particular sections ready for use if the field of data analysis had not evolved to the stage it is now. With the advent of deep learning, the models can themselves decide the structuring, storing, and classification of all the big data and can reproduce it for quick usage as and when required. The complete simplification and handling of all the data is

still not an easy task, but it goes without saying that the technology has still contributed a lot to the development of new drugs, especially the precision medicine.

It would be largely beneficial to the medicinal development if there existed a combined repository of all the medicinal research and data in the world. But the main obstacles in this road are the existence of ethics and commercial viability. The individual patient data is private property and even if it is allowed to be shared then it can be used in a commercial manner. The current nature of the pharmacological industries is a competitive one, and also stands in the way of shared knowledge. There have been certain steps from the professionals around the world to promote the concept in the form of the NCI CCG (Center for Cancer Genomics) promoted GDC (Genomic Data Commons). And the GENIE (Genomics Evidence Neoplasia Information Exchange) created by the AACR (American Association for Cancer Research). These are the ones with the structured data, steps to create repositories of structured data through the conversion of semi-structured and unstructured data are also being taken with the advancement of natural language processing capabilities of the newer

technologies. The significance of this data is unparalleled when it comes to mitigating widespread diseases such as pandemics⁷⁸. The procedures for testing have also been revolutionized through the existence of this technology. Artificial Intelligence has become intelligent enough to now even suggest new methodologies of administration of the medicine and recognize the possible targets for maximum efficiency⁷⁹. GNS Healthcare and Genetech have combined to achieve this purpose. Glaxo Smith Kline has also entered into a similar partnership with Insilico Medicine for the use of the artificial intelligence resources of InsilicoMedicine. A number of other collaborations have occurred for the correct identification of biomarkers⁷⁹⁻⁸². It is difficult to predict the future of this technology precisely but based on the results it has produced, such as advancements in treatments of rare diseases including cystic fibrosis, anemia, and others, it is apparent that these methodologies will become prevalent in the coming years

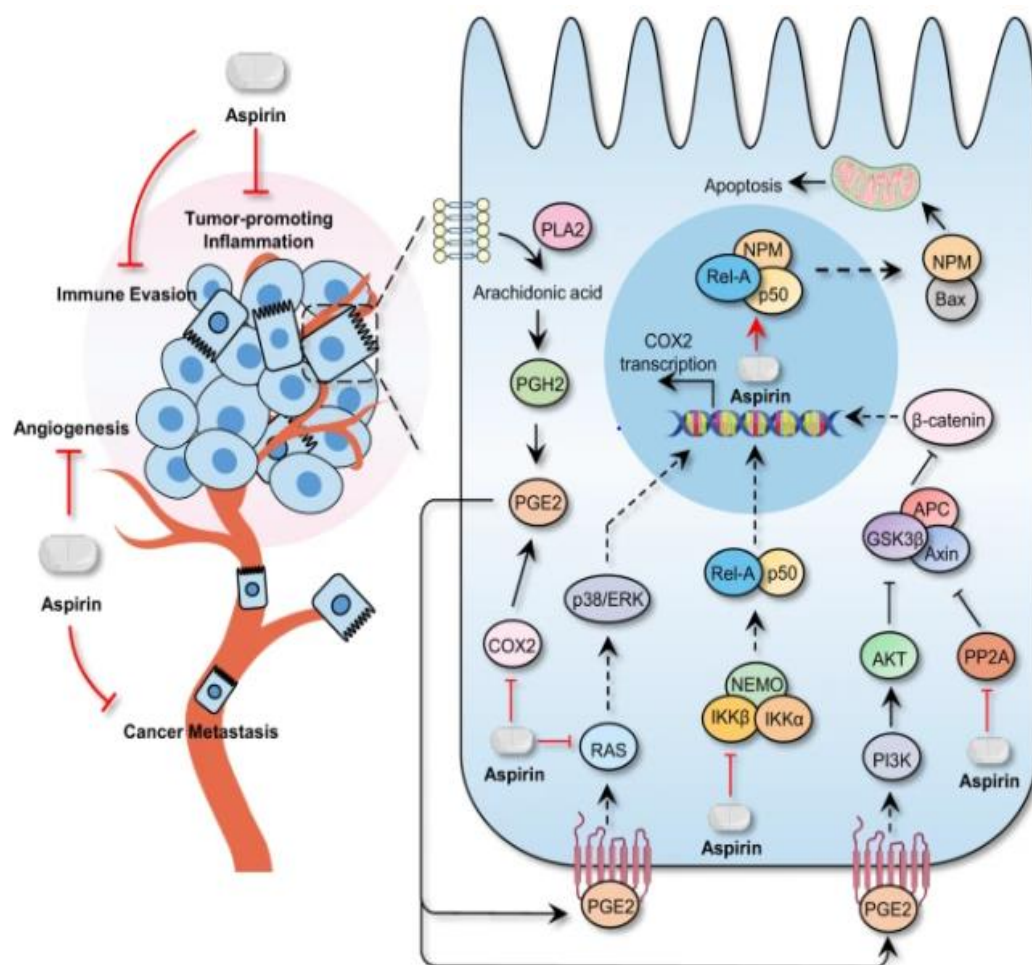


Figure 11: Signaling pathways mediated by metformin. The protein targeting by Metformin is depicted which affects the varied hallmarks of cancer which include regulating cell proliferation, self-renewal, cancer metastasis, angiogenesis and energy metabolism. Reprinted with permission from Springer Nature⁷³, the article is open access and the content reusable. Creative Commons International License 4.0 for the article is available from <http://creativecommons.org/licenses/by/4.0/>

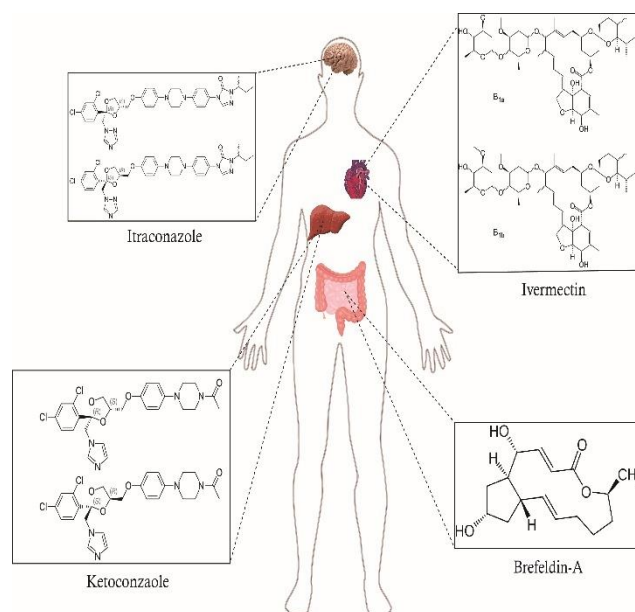


Figure 12: Except for itraconazole, ivermectin, ketoconazole and brefeldin A that are specific to brain, liver, breast and colon cancer, respectively, antifungal agent ciclopirox olamine, anthelmintic drug pyrvinium, and natural products quercetin, 3'-epi-12 β -hydroxyfroside and toxicarioside O have also been studied. Reprinted with permission from Springer Nature⁷³, the article is open access and the content reusable. Creative Commons International License 4.0 for the article is available from <http://creativecommons.org/licenses/by/4.0/>

4.2 AI Integrated Computational Chemistry

The integration of Artificial Intelligence with the field of computational chemistry is also a point of interest to the researchers. Quantum mechanical models can be simplified through AI and when this is combined with molecular dynamics, significant data is revealed. One example of such a situation can be Behler-Parinello's symmetry function that was able to build the potential neural networks for thousands of atoms at the same time¹⁰². A number of applications of this concept exist today^{103,104,113-122,105,123,106-112}. It is a challenging task to integrate artificial intelligence to the sheer randomness of the biological systems. The paramount task to achieve in this field would be to identify placeholders, that is, the points which could efficiently represent a number of data points in their locality. With the development of a process that could yield results to this end, the applications of AI would become infinitely easier. There's only a certain level of extrapolations and predictions that should be made from a certain set of data. With biological data, this limit is difficult to quantify causing the error. The consideration also has to be made about the fact that uncertainty exists at an intrinsic level when the phenomenon of this scale occurs. But even after all that, this technology holds promise and is probable to be widely used in the future.

4.2.1 Active Learning Algorithm Application in Drug Discovery

There are two approaches for using computational chemistry, the structure-based approach and ligand-based approach. One requires information on the ligand only while the other requires explicit knowledge of the ligand and the target. Artificial Intelligence can revolutionize both the methods and provide us with an accurate result. To a further extent, the models with machine learning can outdo all other calculations about quantum mechanics and force fields, provide they have a huge amount of good quality of data to work on

The more common ML methods are the GPR(Gaussian Progress Regression), NN(Neural Network), and KRR(Kernel Ridge Regression); they are replacing the DFT and ab initio method¹²⁴. One of the best points about Artificial Intelligence is that it can work on much larger and many different models once it has been trained with a data-set large enough.

Active learning is a method based on feedback. The conventional computational methods were all based on a set of rigid rules but the feedback loop process combines testing and analyzing to produce a more adaptive computing process that can not only find the best possible outcome but also in each iteration finds the easier methods to realize that outcome, that is, methods with the minimum data required. Built upon this idea of Active Learning is the modified version¹²⁵ of the ANI-1^{126,127}, the ANI-1x; making it not only the best system but also the most economical. An even more accurate model is created by using multiple data sets for training the algorithm. The first data set can be moderately inaccurate but should be large in size, the next one can be smaller but needs to be more accurate. This is done when it is difficult to procure accurate data. It is believed that with the help of large, though less accurate, data it would become easier for the system to find the relations between the smaller, more accurate dataset. The ANI-1ccx¹²⁸ is based upon this methodology.

With the attractive computational powers of these new systems, it is even easier to envision a fully automated method of the calculations of the free energy changes, finding the conformations of the ligands avoiding the false results, novel drug designs, and their synthesis. With the use of robotic techniques, the lineup for automation will be completed. But to do this, the researchers may have to give up on their competitive nature of privacy of their methods, and allow open access so that with combined knowledge from a number of sources, and the ultra-efficient system can be developed. And hopefully, they will be available to surpass the human intuition capabilities as well, and create results human cannot even think to be possible¹²⁴.

4.3 Multiple Target Drug Development using AI

Poly-pharmacology^{129,130} is the idea that prescribes a drug to tackle multiple targets concertedly. Artificial intelligence will play a major role in the recognition of these targets and designing of drugs that would most effectively accomplish this task. A number of major complex diseases can be cured with this approach. Even if a drug is unsuccessful in trials, it may still provide us with valuable information about the existence of a new target site, every time it is introduced in the system. There is no complete description of the interaction that takes place between the medicine and disease at its target. We apply a method known as guilt by association. It involves identifying actions of certain drugs on certain targets only through the action after it has occurred. Machine learning provides us with methods to use

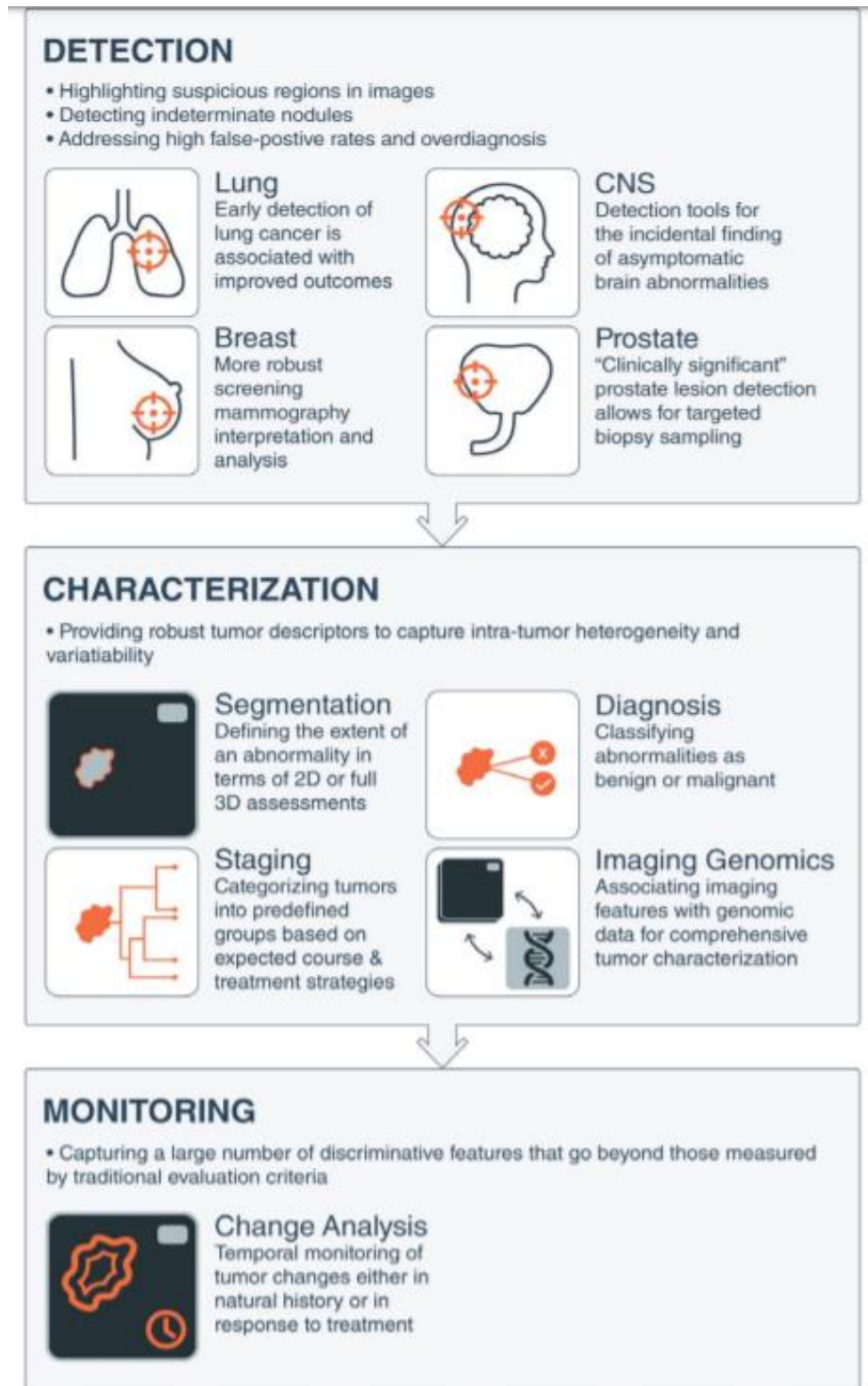


Figure 13: The technology of Machine Learning and Artificial Intelligence can be applied to Cancer Treatment in three major fields of Identification, Characterization, and Monitoring DOI: 10.3322/caac.21552. Reprinted with permissions from the American Cancer Society

the data we have and analyze it to predict the best possible treatments even without the knowledge of the

real mechanisms. Here lies the real power of Artificial intelligence, to predict an effect without an understanding of the cause. The existence of virtual compounds is also of great help to the medicine industry. Through the use of algorithms and models, we can predict, up to a certain extent, the actions of drugs at specific locations at the body and even create simulations as to how they can be optimized for the best results. With the advent of this simulation technology, another possibility has opened up which would help in drug designing. We could just optimize the drug with the help of simulation and then create it, effectively creating a whole new method of precision drug discovery, and also polypharmacological drugs. But there are certain limitations to AI as well, it may have proven to be very effective but in the end, it is still a probability-based concept. There are probable chances of the models not working and instead of causing a negative impact instead. This has to be kept in mind, especially in cases with a low amount of data. The way to get over this would be to integrate computational chemistry methodologies in AI-based models. Quantum Mechanical Calculations can be performed to create extra data, there are other adaptive methods as well. But a combination of seamless generation, accuracy, and simultaneous adaptation is still a theoretical idea. Many pharmaceutical organizations have already begun an attempt to secure this¹³¹.

4.4 Synthesis of Drugs with the use of Artificial Intelligence

The existence of virtual drugs may be helpful in the identification and discovery of new classes of drugs, but the major problem it faces is the fact that every virtual concept cannot be turned into reality. Most of the virtual drugs are either not feasible or have a very discouraging method of synthesis. If the drug is not cost-effective, it is not produced. The identification of the feasible drug is an intuition based-phenomenon as well. Professionals with years of experience and expertise under their belt are the ones with the greatest prospects of being able to create and design those drugs which could be mass-produced¹³². The development of techniques to shift this task to machines is underway. Algorithms such as CAMEO, EROS, LHASA, RASA, SOPHIA, and SYNCHEM have been produced with the same end goal in mind¹³³⁻¹³⁸. Chematica^{139,140} has released a large amount of data on this topic. This task is pretty challenging majorly due to the problem of diversity in the synthesis. There is no single set of rules that is sufficient to dictate the creation of only feasible molecules. To some extent, automation of synthesis^{141,142} is possible but we still haven't reached a stage where it is possible without human intervention.

One method to overcome the lack of a uniform ruling system is to create a ruling system that adapts to the compound. One can argue that through classification, there could exist a finite number of rules systems, but the conditions and variables far exceed designing capability. But a method to bypass this is theoretically possible in algorithms based on learning. We may provide the algorithm with reactions and allow it to learn on its own through observations of the reactions. This type of modelling might yield accurate variable ruling systems for different types of synthesis. The machine does not identify any real mechanisms for this job but is simply dependent on components that vary in different situations. An accurate prediction is possible with them. These types of experiments have already been performed and some encouraging results have been received^{143,144}. These systems can also use chemical data such as graphs to teach the algorithm into becoming an efficient producer¹⁴⁵. An even more efficient method includes synthesis feasibility at a number of steps. The code keeps on comparing the expenses and availability of the necessary conditions and raw materials that would allow the product to exist in reality. This type of code saves a lot of computing power and time. Another approach is to also include the negatively resulting arguments into the system. These negative reactions would allow for the easy elimination of unreasonable products.

As a whole, there is a lot to be explored in the domain of machine learning and artificial intelligence for drug discovery and synthesis. The fact that AI can read through the humongous quantity of texts, choose the best possible combination for design, find the

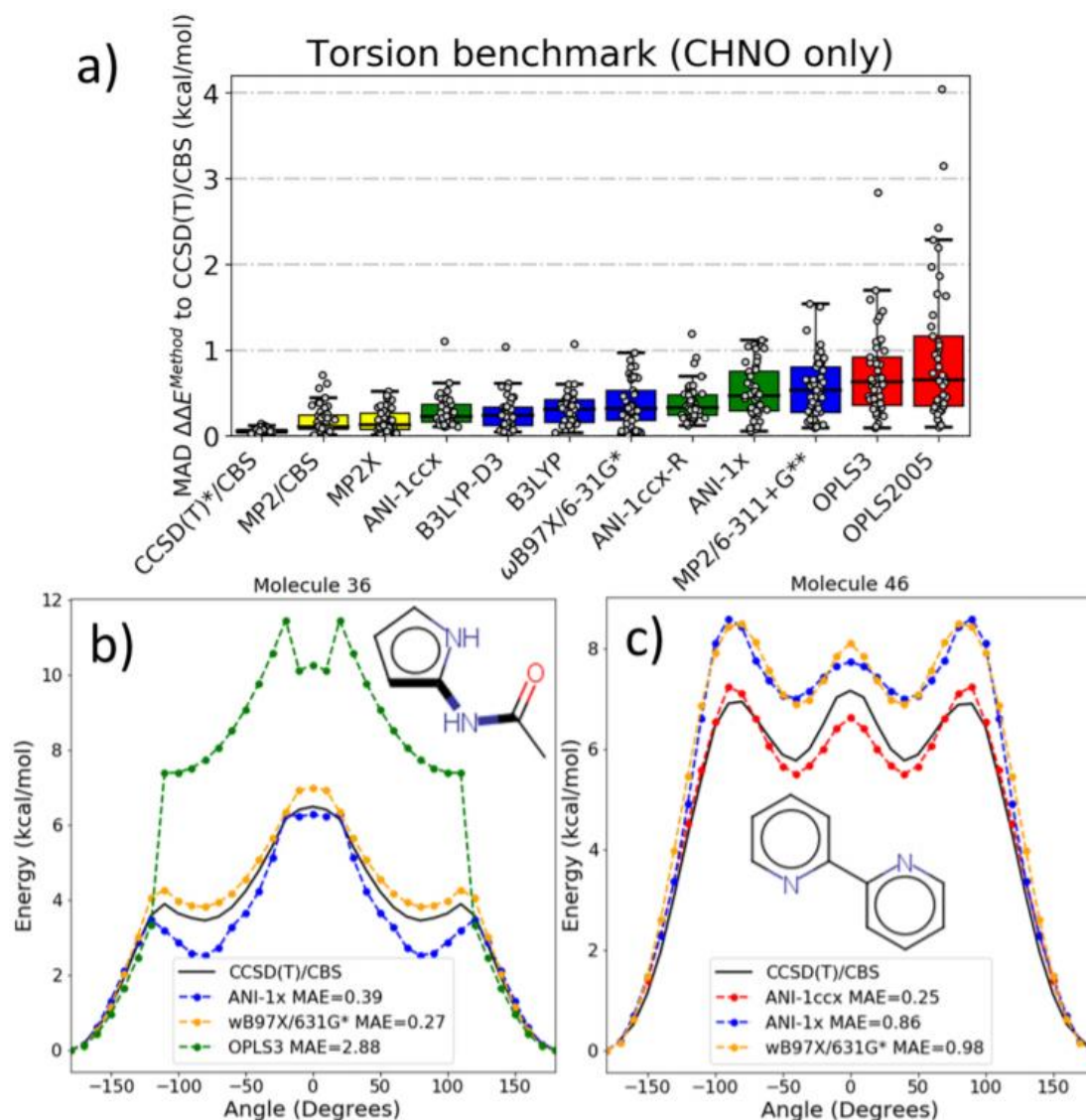


Figure 14: Computational Chemistry model ANI integrated with machine learning and data analysis technology turned into ANI-1xReprinted (adapted) with permission from American Chemical Society

optimal route of administration and even find the best method to synthesize the required drug; and the fact that this could be the lifetime of work of a chemist and a machine can perform it within days makes the field of Artificial

Intelligence a very desirable and productive field when it comes to pharmacology^{144,145}.

5. Refurbishment of Global Health with AI

Low middle-income countries (LMICs) have always suffered due to poor health care systems owing to low investments, inadequate training providers, and ease of access to the latest developments in medical sciences to improve their health care system. AI has shown great opportunities to bolster and ameliorate the existing health care in LMICs. AI utilizes available datasets to analyze and further predict with the use of advanced machine learning algorithms and intelligent computing resources. The Neural Networks along with machine learning attributing to their capabilities of learning from experiments and undergo modifications required as per the situation have empowered the use of AI in each aspect. Consequently, it has led to the announcements by various countries for the incorporation of AI in the health care sector for its alleviation and betterment.

But there are still challenges ahead for AI implementation and application for maximum efforts. AI majorly relies on the availability of large datasets to get trained and produce accurate results. Thus, a limited dataset availability in these LMICs is a big challenge to overcome. Recently the data circulated is from high-income settings with not much testing and approved validation in low-income settings. This kind of dataset can cause biases for the training of AI models. Further, there is a compulsory sophistication required during regulation and oversight for these AI models to adapt and learn. Carelessness can cause severe harm greater than any other technology. AI also faces challenges for its implementation in health care. Firstly, they are required to be validated and used by the right set of workers. Then there is a challenge to train the undertrained health workers in LMICs. Lastly, the implementation of AI comes with a secure and effective system to regulate it. In the case of LMICs, even traditional systems are underregulated thus these are some points we need to combat to achieve the successful use of AI in the medical sector.

However, AI is overcoming all these challenges and is becoming an integral part of many leading pharmaceuticals. For instance, Sanofi has joined hands with Exscientia's artificial intelligence platform for devising therapies related to metabolic-diseases. An immuno-oncology drug searching system is now being used by Pfizer. GNS Healthcare in Cambridge is providing AI platforms to Roche subsidiary Genentech to bolster a Multi-national company's search drive for cancer treatments. These advancements are indicating a bright, cost-effective, and highly effective AI-driven era in the health care sector. Even big market support is now coming together for AI and its implementation.

Conclusions

As prevalent everywhere, Artificial Intelligence technologies are on the quest of replacing human intelligence to a great extent. The development of AI algorithms, machine learning, etc. has facilitated this advancement. Another factor adding to the growing prominence of AI technologies is the availability of volumes of data and highly developed computational systems. AI has surpassed the expectations of the current generation with its applications visible everywhere, from image recognition technologies to the envisioning of driverless vehicles. After all these fascinating achievements it is really believable that AI could help the pharmaceutical industry as well by making operations like drug discovery, drug development, drug repurposing, etc. more efficient and easier than the conventional method. Though AI is considered to be an extremely helpful technology but there are certain challenges associated with it as well like the unavailability of high-quality data, the high costs incurred in collecting or testing the data acquired due to which the AI application becomes economically unaffordable. In the study of a drug, the amount of data that are available are way too small compared to the volumes needed for the development of a model. As previously mentioned, the quality of data is a major concern and one of the solutions for this is to create algorithms that have the capacity to bridge the gaps caused due to the heterogeneity in the data available. Models like the Transfer model have been considered worthy of countering these issues. There are thousands of diseases that have no treatment, however, the availability of genomic data and developed computational systems, AI platforms have been developed to repurpose the existing drugs so that a possible treatment for these diseases can be found. There are several tools, both supervised and unsupervised, which have been developed in order to repurpose the drugs. While treating a patient suffering from a disease, doctors try to treat them with a combination of drugs that is best suited to their body's capacity. It is well known to the medical researchers that for each patient, that combination would be different, and to treat a disease it is important that a proper combination is concocted. The task is too exacting and thus AI platforms like QPOP are used to delineate a large number of available drugs and find the perfect therapeutic combination. Diseases like cancer are unique in the fact that their treatment requires individual analysis of the condition of each patient and that yields a large amount of data. The technologies like Artificial Intelligence and Machine Learning are employed here as well to process the volumes of data and provide a personalized analysis.

Conflicts of interest

There are no conflicts to declare

Notes and references

1. 1 J. N. Weinstein, K. W. Kohn, M. R. Grever, V. N. Viswanadhan, L. V Rubinstein, A. P. Monks, D. A. Scudiero, L. Welch, A. D. Koutsoukos, A. J. Chiausa and al. et, *Science (80-.)*, 1992, **258**, 447–451.
2. 2 G. Schneider, *Nat. Mach. Intell.*, 2019, **1**, 128.
3. 3 L. Yu, S. Wang and K. K. Lai, *IEEE Trans. Knowl. Data Eng.*, 2006, **18**, 217.
4. 4 R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, 2008.
5. 5 Y.-C. Lo, S. E. Rensi, W. Torng and R. B. Altman, *Drug Discov. Today*, 2018, **23**, 1538.
6. 6 R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics: Vol. I: Alphabetical Listing/Vol. II: Appendices, References*, 2009.
7. 7 T. Unterthiner, A. Mayr, G. Klambauer, M. Steijaert, J. K. Wegner, H. Ceulemans and S. Hochreiter, *Proceedings of the Deep Learning Workshop at NIPS*, 2014.
8. 8 T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath and H. Chen, *Mol. Inf.*, 2018, **37**, 1700123.
9. 9 N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *J. Mach. Learn. Res.*, 2014, **15**, 1929.
10. 10 C. Rücker, G. Rücker and M. Meringer, *J. Chem. Inf. Model.*, 2007, **47**, 2345.
11. 11 F. Provost, *Proceedings of the AAAI'2000 Workshop on Imbalanced Data Sets*, 2000.
12. 12 C. E. Brodley and M. A. Friedl, *J. Artif. Intell. Res.*, 1999, **11**, 131.
13. 13 J. Van Hulse and T. Khoshgoftaar, *Data Knowl. Eng.*, 2009, **68**, 1513.
14. 14 N. V Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, *J. Artif. Intell. Res.*, 2002, **16**, 321.
15. 15 T. Maciejewski and J. Stefanowski, *2011 IEEE Symposium on Computational Intelligence and Data Mining*, 2011.
16. 16 X. Xiao, J.-L. Min, W.-Z. Lin, Z. Liu, X. Cheng and K.-C. Chou, *J. Biomol. Struct. Dyn.*, 2015, **33**, 2221.
17. 17 H. Altae-Tran, B. Ramsundar, A. S. Pappu and V. Pande, *ACS Cent. Sci.*, 2017, **3**, 283.
18. 18 A. V Zakharov, M. L. Peach, M. Sitzmann and M. C. Nicklaus, *J. Chem. Inf. Model.*, 2014, **54**, 705.
19. 19 R. Guha and S. C. Schürer, *J. Comput. Mol. Des.*, 2008, **22**, 367.
20. 20 D. C. Weis, D. P. Visco and J.-L. Faulon, *J. Mol. Graph. Model.*, 2008, **27**, 466.
21. 21 G. V Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason and A. L. Hopkins, *Nat. Biotechnol.*, 2006, **24**, 805.
22. 22 U. Koch, M. Hamacher and P. Nussbaumer, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 156.
23. 23 B. S. Glicksberg, L. Li, W.-Y. Cheng, K. Shameer, J. Hakenberg, R. Castellanos, M. Ma, L. Shi, H. Shah and J. T. Dudley, *Pacific Symposium on Biocomputing 2015*, 2014.
24. 24 R. M. Piro, *Hum. Genet.*, 2012, **131**, 1811.
25. 25 T. T. Ashburn and K. B. Thor, *Nat. Rev. Drug Discov.*, 2004, **3**, 673.
26. 26 J. T. Dudley, T. Deshpande and A. J. Butte, *Briefings Bioinf.*, 2011, **12**, 303.
27. 27 M. Hurlle, L. Yang, Q. Xie, D. Rajpal, P. Sansseau and P. Agarwal, *Clin. Pharmacol. Ther.*, 2013, **93**, 335.
28. 28 D. Sardana, C. Zhu, M. Zhang, R. C. Gudivada, L. Yang and A. G. Jegga, *Briefings Bioinf.*, 2011, **12**, 346.
29. 29 V. J. Haupt and M. Schroeder, *Briefings Bioinf.*, 2011, **12**, 312.
30. 30 E. De Franchi, C. Schalon, M. Messa, F. Onofri, F. Benfenati and D. Rognan, *PLoS One*, 2010, **5**, e12214.
31. 31 S. L. Kinnings, N. Liu, N. Buchmeier, P. J. Tonge, L. Xie and P. E. Bourne, *PLoS Comput. Biol.*, 2009, **5**, e1000423.
32. 32 M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijter, R. C. Matos and T. B. Tran, *Nature*, 2009, **462**, 175.
33. 33 E. C. Butcher, E. L. Berg and E. J. Kunkel, *Nat. Biotechnol.*, 2004, **22**, 1253.
34. 34 R. Bellazzi, M. Diomidous, I. N. Sarkar, K. Takabayashi, A. Ziegler and A. T. McCray, *Methods Inf. Med.*, 2011, **50**, 536.
35. 35 R. Margolis, L. Derr, M. Dunn, M. Huerta, J. Larkin, J. Sheehan, M. Guyer and E. D. Green, *J. Am. Med. Inform. Assoc.*, 2014, **21**, 957.
36. 36 A. Gottlieb, G. Y. Stein, E. Ruppel and R. Sharan, *Mol. Syst. Biol.*, 2011, **7**, 496.
37. 37 Z. Xia, L.-Y. Wu, X. Zhou and S. T. Wong, *BMC Syst. Biol.*, 2010, **4**, S6.
38. 38 Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen and J. Zeng, *Nat. Commun.*, 2017, **8**, 573.
39. 39 F. Napolitano, Y. Zhao, V. M. Moreira, R. Tagliaferri, J. Kere, M. D'Amato and D. Greco, *J. Cheminformatics.*, 2013, **5**, 30.
40. 40 M. Oh, J. Ahn and Y. Yoon, *PLoS One*, 2014, **9**, e111668.
41. 41 Z. Liu, F. Guo, J. Gu, Y. Wang, Y. Li, D. Wang, L. Lu, D. Li and F. He, *Bioinformatics*, 2015, **31**, 1788.
42. 42 W. Dai, X. Liu, Y. Gao, L. Chen, J. Song, D. Chen, K. Gao, Y. Jiang, Y. Yang and J. Chen, *Comput. Math. Methods Med.*, 2015, **2015**, 275045.
43. 43 H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu and Y. Pan, *Bioinformatics*, 2016, **32**, 2664.
44. 44 F. Napolitano, Y. Zhao, V. M. Moreira, R. Tagliaferri, J. Kere, M. D'Amato and D. Greco, *J. Cheminform.*, 2013, **5**, 30.
45. 45 T. Noeske, B. C. Sasse, H. Stark, C. G. Parsons, T. Weil and G. Schneider, *ChemMedChem*, 2006, **1**, 1066.
46. 46 D. Carrella, F. Napolitano, R. Rispoli, M. Miglietta, A. Carissimo, L. Cuttillo, F. Sirici, F. Gregoretti and D. Di Bernardo, *Bioinformatics*, 2014, **30**, 1787.
47. 47 B. J. Frey and D. Dueck, *Science (80-.)*, 2007, **315**, 972.
48. 48 Z. Wu, X.-M. Zhao and L. Chen, *BMC Syst. Biol.*, 2010, **4**, S7.
49. 49 M. Belkin, P. Niyogi and V. Sindhwani, *J. Mach. Learn. Res.*, 2006, **7**, 2399.
50. 50 X.-Y. Yan, S.-W. Zhang and S.-Y. Zhang, *Mol. BioSyst.*, 2016, **12**, 520.
51. 51 J.-P. Mei, C.-K. Kwok, P. Yang, X.-L. Li and J. Zheng, *Bioinformatics*, 2013, **29**, 238.

52. 52 Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436.
53. 53 A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina and A. Zhavoronkov, *Mol. Pharm.*, 2016, **13**, 2524–2530.
54. 54 Z. M. Hira and D. F. Gillies, *Adv. Bioinf.*, 2015, **2015**, 198363.
55. 55 P. V Spirin, T. D. Lebedev, N. N. Orlova, A. S. Gornostaeva, M. M. Prokofjeva, N. A. Nikitenko, S. E. Dmitriev, A. A. Buzdin, N. M. Borisov, A. M. Aliper, A. V Garazha, P. M. Rubtsov, C. Stocking and V. S. Prassolov, *Leukemia*, 2014, **28**, 2222.
56. 56 Q. Zhu, E. Izumchenko, A. M. Aliper, E. Makarev, K. Paz, A. A. Buzdin, A. A. Zhavoronkov and D. Sidransky, *Hum. Genome Var.*, 2015, **2**, 15009.
57. 57 A. A. Buzdin, A. A. Zhavoronkov, M. B. Korzinkin, L. S. Venkova, A. A. Zenin, P. Y. Smirnov and N. M. Borisov, *Front. Genet.*, 2014, **5**, 55.
58. 58 A. Artemov, A. Aliper, M. Korzinkin, K. Lezhnina, L. Jellen, N. Zhukov, S. Roumiantsev, N. Gaifullin, A. Zhavoronkov, N. Borisov and A. Buzdin, *Oncotarget*, 2015, **6**, 29347.
59. 59 L. Venkova, A. Aliper, M. Suntsova, R. Kholodenko, D. Shepelin, N. Borisov, G. Malakhova, R. Vasilov, S. Roumiantsev, A. Zhavoronkov and A. Buzdin, *Oncotarget*, 2015, **6**, 27227.
60. 60 E. Makarev, C. Cantor, A. Zhavoronkov, A. Buzdin, A. Aliper and A. B. Csoka, *Aging (Albany. NY.)*, 2014, **6**, 1064.
61. 61 C. J. Langmead, J. Watson and C. Reavill, *Pharmacol. Ther.*, 2008, **117**, 232.
62. 62 L. A. Volpicelli and A. I. Levey, *Prog. Brain Res.*, 2004, **145**, 59.
63. 63 J. S. Trimmer, *Neuron*, 2015, **85**, 238.
64. 64 T. Li, H. J. Kung, P. C. Mack and D. R. Gandara, *J. Clin. Oncol.*
65. 65 C. Pauli, *Cancer Discov.*
66. 66 A. J. Aguirre, *Cancer Discov.*
67. 67 C. Rubio-Perez, *Cancer Cell.*
68. 68 F. Cheng, *Nat. Commun.*
69. 69 M. N. Patel, *Nat. Rev. Drug Discov.*
70. 70 A. B. Sarmiento-Ribeiro, *Drug Resist Updat.*
71. 71 A. Boyer, *Eur. Respir. Rev.*
72. 72 Q. I. Zhang, *Oncol. Lett.*
73. 73 Z. Zhang, L. Zhou, N. Xie, E. C. Nice, T. Zhang, Y. Cui and C. Huang, *Signal Transduct. Target. Ther.*, 2020, **5**, 113.
74. 74 D. Hanahan and R. A. Weinberg, *Cell.*
75. 75 L. Sleire, *Pharm. Res.*
76. 76 Y. Chen, *J. Hepatol.*
77. 77 S. A. Dugger, A. Platt and D. B. Goldstein, *Nat. Rev. Drug Discov.*, 2018, **17**, 183.
78. 78 D. A. Broniatowski, M. J. Paul and M. Dredze, *PLoS One*, 2013, **8**, e83672.
79. 79 B. Chen and A. Butte, *Clin. Pharmacol. Ther.*, 2016, **99**, 285.
80. 80 T. Zeng, W. Zhang, X. Yu, X. Liu, M. Li and L. Chen, *Briefings Bioinf.*, 2016, **17**, 576.
81. 81 W.-S. Lee, S.-M. Ahn, L. Uhn, N. Skjodt and A. Zhavoronkov, *J. Gerontol A Biol. Sci. Med. Sci.*, 2018, **73**, 1482.
82. 82 D. W. Belsky, K. M. Huffman, C. F. Pieper, I. Shalev and W. E. Kraus, *J. Gerontol., Ser. A*, 2018, **73**, 4.
83. 83 R. A. Castellino, *Cancer Imaging*, 2005, **5**, 17–19.
84. 84 M. Liang, W. Tang, D. M. Xu, A. C. Jirapatnakul, A. P. Reeves, C. I. Henschke and D. Yankelevitz, *Radiology*, 2016, **281**, 279–288.
85. 85 J. M. Lewin, ed. M. A. B. T.-C. I. Hayat, Academic Press, San Diego, 2008, pp. 455–458.
86. 86 T. S. Hong, W. A. Tomé and P. M. Harari, *Radiother. Oncol.*, 2012, **103**, 92–98.
87. 87 X. A. Li, A. Tai, D. W. Arthur, T. A. Buchholz, S. Macdonald, L. B. Marks, J. M. Moran, L. J. Pierce, R. Rabinovitch, A. Taghian, F. Vicini, W. Woodward and J. R. White, *Int. J. Radiat. Oncol. Biol. Phys.*, 2009, **73**, 944–951.
88. 88 S. K. Warfield, K. H. Zou and W. M. Wells, *Philos. Trans. Ser. A, Math. Phys. Eng. Sci.*, 2008, **366**, 2361–2375.
89. 89 H.-P. Chan, L. Hadjiiski, C. Zhou and B. Sahiner, *Acad. Radiol.*, 2008, **15**, 535–555.
90. 90 S. Wang, K. Burt, B. Turkbey, P. Choyke and R. M. Summers, *Biomed Res. Int.*, 2014, **2014**, 789561.
91. 91 C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk and J. V Lebesque, *Int. J. Radiat. Oncol.*, 1999, **43**, 57–66.
92. 92 O. Ruprecht, P. Weisser, B. Bodelle, H. Ackermann and T. J. Vogl, *Eur. J. Radiol.*, 2012, **81**, 456–460.
93. 93 S. Mirsadraee, D. Oswal, Y. Alizadeh, A. Caulo and E. J. van Beek, *World J. Radiol.*, 2012, **4**, 128–134.
94. 94 M. B. Amin, F. L. Greene, S. B. Edge, C. C. Compton, J. E. Gershenwald, R. K. Brookland, L. Meyer, D. M. Gress, D. R. Byrd and D. P. Winchester, *CA. Cancer J. Clin.*, 2017, **67**, 93–99.
95. 95 G. Levrini, R. Sghedoni, C. Mori, A. Botti, R. Vacondio, A. Nitrosi, M. Iori and F. Nicoli, *Radiol. Med.*, 2011, **116**, 1039–1049.
96. 96 C. Parmar, J. D. Barry, A. Hosny, J. Quackenbush and H. J. W. L. Aerts, *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.*, 2018, **24**, 3492–3499.
97. 97 C. C. Jaffe, *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*, 2006, **24**, 3245–3251.
98. 98 J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird and N. Rosenfeld, *Nat. Rev. Cancer*, 2017, **17**, 223–238.
99. 99 C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan,

- A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann and C. Swanton, *Nature*, 2017, **545**, 446–451.
100. 100 J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M.-B. W. Orntoft, B. D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamalay, A. Georgiadis, Q. K. Li, M. R. Madsen, F. V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R. B. Scharpf, L. A. Diaz, S. Jones, S. Angiuoli, T. Ørntoft, H. J. Nielsen, C. L. Andersen and V. E. Velculescu, *Sci. Transl. Med.*, 2017, **9**, eaan2415.
101. 101 J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbryn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. J. Diaz, C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon and N. Papadopoulos, *Science*, 2018, **359**, 926–930.
102. 102 J. Behler, *J. Chem. Phys.*, 2016, **145**, 170901.
103. 103 L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U. N. Niranjana, P. Duncan, M. Rupp, K. R. Müller and K. Burke, *Int. J. Quantum Chem.*, 2016, **116**, 819.
104. 104 J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller and K. Burke, *Phys. Rev. Lett.*, 2012, **108**, 253002.
105. 105 F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke and K.-R. Müller, *Nat. Commun.*, 2017, **8**, 872.
106. 106 K. Yao and J. Parkhill, *J. Chem. Theory Comput.*, 2016, **12**, 1139.
107. 107 J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller and K. Burke, *J. Chem. Phys.*, 2013, **139**, 224104.
108. 108 K. Mills, M. Spanner and I. Tamblin, *Phys. Rev. A At., Mol., Opt. Phys.*, 2017, **96**, 42113.
109. 109 F. Liu, L. Du, D. Zhang and J. Gao, *Sci. Rep.*, 2017, **7**, 8737.
110. 110 F. Häse, S. Valleau, E. Pyzer-Knapp and A. Aspuru-Guzik, *Chem. Sci.*, 2016, **7**, 5139.
111. 111 B. K. Carpenter, G. S. Ezra, S. C. Farantos, Z. C. Kramer and S. Wiggins, *J. Phys. Chem. B*, 2018, **122**, 3230.
112. 112 K. Yao, J. E. Herr and J. Parkhill, *J. Chem. Phys.*, 2017, **146**, 14106.
113. 113 Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 96405.
114. 114 V. Botu and R. Ramprasad, *Int. J. Quantum Chem.*, 2015, **115**, 1074.
115. 115 G. Pilania, A. Mannodi-Kanakkithodi, B. Uberuaga, R. Ramprasad, J. Gubernatis and T. Lookman, *Sci. Rep.*, 2016, **6**, 19375.
116. 116 G. Pilania, J. E. Gubernatis and T. Lookman, *Comput. Mater. Sci.*, 2017, **129**, 156.
117. 117 A. Mannodi-Kanakkithodi, G. Pilania, T. D. Huan, T. Lookman and R. Ramprasad, *Sci. Rep.*, 2016, **6**, 20952.
118. 118 T. D. Huan, A. Mannodi-Kanakkithodi and R. Ramprasad, *Phys. Rev. B Condens. Matter Mater. Phys.*, 2015, **92**, 14106.
119. 119 G. Pilania, C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, *Sci. Rep.*, 2013, **3**, 2810.
120. 120 J. Lee, A. Seko, K. Shitara, K. Nakayama and I. Tanaka, *Phys. Rev. B Condens. Matter Mater. Phys.*, 2016, **93**, 115104.
121. 121 E. O. Pyzer-Knapp, K. Li and A. Aspuru-Guzik, *Adv. Funct. Mater.*, 2015, **25**, 6495.
122. 122 R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha and T. Wu, *Nat. Mater.*, 2016, **15**, 1120.
123. 123 X. Ma, Z. Li, L. E. Achenie and H. Xin, *J. Phys. Chem. Lett.*, 2015, **6**, 3528.
124. 124 J. S. Smith, A. E. Roitberg and O. Isayev, *ACS Med. Chem. Lett.*, 2018, **9**, 1065–1069.
125. 125 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. A. E. Roitberg, *J. Chem. Phys.*, 2018, **148**, 241733.
126. 126 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192.
127. 127 J. S. Smith, O. Isayev and A. E. Roitberg, *Sci. Data*, 2017, **4**, 170193.
128. 128 J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. Roitberg, *ChemRxiv*, 2018.
129. 129 A. Anighoro, J. Bajorath and G. Rastelli, *J. Med. Chem.*, 2014, **57**, 7874.
130. 130 W. Zhang, J. Pei and L. Lai, *J. Chem. Inf. Model.*, 2017, **57**, 403.
131. 131 X. Yang, Y. Wang, R. Byrne, G. Schneider and S. Yang, *Chem. Rev.*, 2019, **119**, 10520–10594.
132. 132 K. D. Collins and F. Glorius, *Nat. Chem.*, 2013, **5**, 597.
133. 133 W. L. Jorgensen, E. R. Laird, A. J. Gushurst, J. M. Fleischer, S. A. Gothe, H. E. Helson, G. D. Paderes and S. Sinclair, *Pure Appl. Chem.*, 1990, **62**, 1921.
134. 134 P. Röse and J. Gasteiger, *Software Development in Chemistry 4*, 1990.
135. 135 E. Corey, *Q. Rev., Chem. Soc.*, 1971, **25**, 455.
136. 136 Q. Huang, L.-L. Li and S.-Y. Yang, *J. Chem. Inf. Model.*, 2011, **51**, 2768.
137. 137 H. Satoh and K. Funatsu, *J. Chem. Inf. Model.*, 1995, **35**, 34.
138. 138 H. Gelernter, A. Sanders, D. Larsen, K. Agarwal, R. Boivie, G. Spritzer and J. Searleman, *Science (80-)*, 1977, **197**, 1041.
139. 139 M. Fialkowski, K. J. Bishop, V. A. Chubukov, C. J. Campbell and B. A. Grzybowski, *Angew. Chem., Int. Ed.*, 2005, **44**, 7263.

-
140. 140 S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk and B. A. Grzybowski, *Angew. Chem., Int. Ed.*, 2016, **55**, 5904.
141. 141 D. Fooshee, A. Andronico and P. Baldi, *J. Chem. Inf. Model.*, 2013, **53**, 2812.
142. 142 N. Schneider, N. Stiefl and G. A. Landrum, *J. Chem. Inf. Model.*, 2016, **56**, 2336.
143. 143 J. N. Wei, D. Duvenaud and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2016, **2**, 725.
144. 144 C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green and K. F. Jensen, *ACS Cent. Sci.*, 2017, **3**, 434.
145. 145 M. H. Segler and M. P. Waller, *Chem. - Eur. J.*, 2017, **23**, 6118.
146. 146 F. Iorio, R. Bosotti, E. Scacheri, V. Belcastro, P. Mithbaokar, R. Ferriero, L. Murino, R. Tagliaferri, N. Brunetti-Pierri and A. Isacchi, *P. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14621.
147. 147 D. Ho, *Science (80-.)*, 2020, **367**, 982–983.