Chapter 1
The Application of AI in Precision Oncology: Tailoring Diagnosis, Treatment, and the Monitoring of Disease Progression to the Patient

Zodwa Dlamini and Rodney Hull

Abstract Personalised oncology has long been the ideal when it comes to the management of cancer. The ability to tailor screening, diagnosis, therapy and monitoring to an individual patient or group of patients would vastly decrease the burden of cancer while ensuring higher rates of patient survival and treatments with less side effects and more success in controlling or eliminating the disease. Precision oncology requires that as much information regarding the patient or population group be known. In terms of the underlying molecular basis of the disease, this is now being realised further to the advent of high throughput technologies such as next-generation sequencing (NGS) and advances in mass spectrophotometry. This has led to an "omics" revolution, with large datasets of information regarding the molecular basis of cancer in individuals being generated. Artificial intelligence (AI) is the ideal technology to manage and interpret these large datasets. In conjunction with machine learning (ML) and deep learning (DL), AI can more accurately interpret not only omics data, but it can also integrate data from other sources such as patient reports and medical imaging to give a more precise view of the individual or population, allowing for better clinical decision-making.

Z. Dlamini (✉) · R. Hull
SAMRC Precision Oncology Research Unit (PORU), DSI/NRF SARChI Chair in Precision Oncology and Cancer Prevention (POCP), Pan African Cancer Research Institute (PACRI), University of Pretoria, Pretoria, South Africa
e-mail: zodwa.dlamini@up.ac.za; rodney.hull@up.ac.za

1.1 Introduction

The underlying molecular basis of cancer is complex and deciphering it has been the basis for many decades of research. The revolution in the available techniques that occurred in the 1970s led to the first in-depth studies concerning the molecular basis of cancer. Gaining an understanding of the changes that give rise to cancer at the molecular level allowed to not only understand how events in the body give rise to the disease but also how it progresses and also how these events could be targeted for the development of therapies. It was these initial studies that allowed for the development of the first drugs that could target molecules and signalling pathways to treat oncogenic processes, such as uncontrolled proliferation and resistance to apoptosis. When microarrays were first developed, their ability to create a profile of gene transcription in cancer (reviewed in (Govindarajan et al., 2012)) led to the first discussions of precision oncology. Precision oncology, a type of precision medicine, involves the tailoring of screening or treatment to an individual or specific population group based on the molecular profiles specific to that individual or group of individuals (Batch et al., 2022). The understanding of the molecular biology underlying cancer has been advanced in recent decades by the development of high throughput techniques such as next-generation sequencing (NGS) and advanced proteomics techniques such as SWATH. The data generated by these techniques has been used to decipher the molecular mechanisms of tumour initiation and progression. This data has also been used to construct database resources to integrate and analyse molecular mechanisms underlying cancer.

The ability of scientists to use these large datasets and databases to make useful observations and predictions concerning cancer is due to the advent and application of artificial intelligence. Artificial intelligence (AI) is an analytical or predictive operation performed by computers to emulate the decision-making processes of human beings. It has intensive problem-solving capabilities and can be used to perform tasks such as making predictions, scaling data, integrating different datasets, and reducing the dimensionality of data. Most importantly precision oncology can associate different patterns within data with real-world diagnoses, prognoses, or disease monitoring capabilities. The ability of AI to analyse large sets of data and transform this data into clinically actionable knowledge relies on the ability of AI to learn from either previous data or model teaching datasets. This learning ability is based on machine learning (ML) and deep learning (DL)-based approaches (Jiang et al., 2017) (Saltz et al., 2018) (Huang et al., 2020) (Ibrahim et al., 2020). The increase in interest in AI, precision oncology and precision medicine can be seen in the number of entries these topics find when used as search terms in PubMed. Standalone terms that carry entries for AI or precision medicine go back to the 1950s while the earliest entries for precision oncology date to the 1970s. There has also been a lot of interest in AI since the 1990s while interest in the other two increased rapidly from 2010 onwards. A combination of terms involving AI both precision medicine and precision oncology AND AI have only been topics of interest since 2017 (Fig. 1.1).
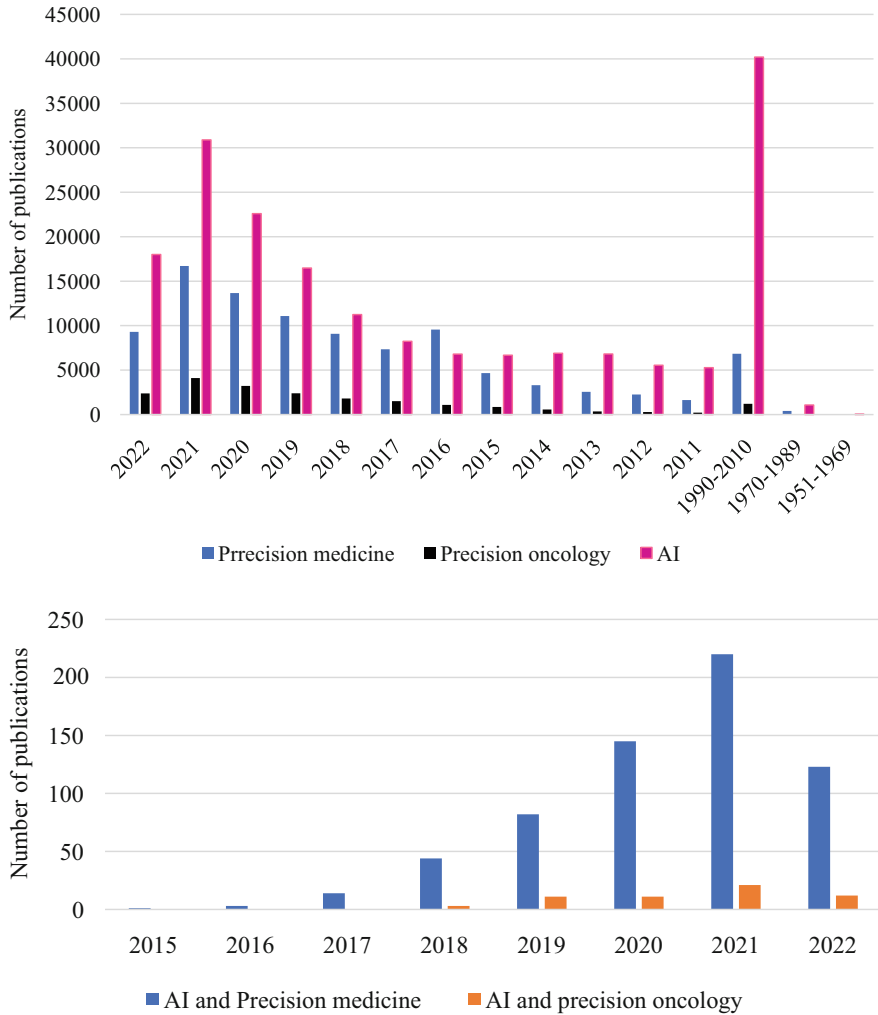
Fig. 1.1 PubMed entries on AI and precision medicine/oncology (**a**) using the terms independently the earliest references to AI or precision medicine come from the 1950s. While the earliest reference to precision oncology comes from 1977. All terms show an increase in the number of entries in PubMed. (**b**) The number of entries in PubMed for AI AND precision medicine and AI AND precision oncology since 2015/ This is a depiction of the number of entries identified in PubMed when the search terms AI AND precision medicine and AI AND precision oncology are used. For both terms, there are only regular entries after 2015 and the number of papers increases dramatically as time goes on. This demonstrates that these are topics of growing interest to researchers

1.2   AI in Medicine

In order for AI to accurately make predictions regarding a patient's health and treatment requirements, it must be able to learn from previous data and analyses. In this way, it emulates the human clinician learning from past experiences. The ability of computing algorithms to learn and adjust their performance to better recognise patterns in data is known as machine learning (ML). Initially, an AI does this using training data to create or fine tune mathematical models (Hakenberg et al., 2012). Deep learning (DL) is a specific type of ML which uses data that is labelled (supervised) and data that is unlabeled (unsupervised) in the training process. It integrates these different types of data by using multi-layer non-linear analysis and classification. One of the applications of DL is in a process known as natural language processing, and reinforcement learning (Falk et al., 2019) (Kaelbling et al., 1996).

Natural language processing (NLP) algorithms use two terms and establish if they are linked by counting the number of times they occur together. If they occur together more frequently than they are associated (Cheng et al., 2008) (Santus et al., 2019). This technique is used to search large amounts of literature or databases of information for articles or cases of interest. This is important because there are vast amounts of literature relating to cancer research and studies. One of these algorithms, known as MEDscape uses NLP to search and organise medical patient notes. The useful data retrieved from these notes is used to automatically update patient records (Morin et al., 2021). AI using NLP algorithms have been used to accurately predict patient outcomes using a variety of data including imaging reports and oncologist notes from thousands of patients with multiple different tumour types. The predictions the AI was able to make included cancer progression, treatment response and the likelihood and speed of metastasis (Kehl et al., 2021).

AI makes use of neural networks to copy the way humans think and interpret data but without user bias and human error. These neural networks allow AI to make logical conclusions similar to those that could be reached by humans (Joshi et al., 2021). These networks use multiple fundamental computing units (neurons) to convert raw input data into classified, annotated and analysed output data. The nodes are connected to form a network that contains multiple layers including an input layer, multiple functional or hidden layers, and an output layer (Kuwahara et al., 2021). There are multiple types of neural networks. Artificial Neural Networks (ANNs) use multiple interconnected computational neurons that distribute data analysis tasks. These networks are useful for analysing multidimensional complex data. The distribution and initial decisions the network make regarding this data are based on the learning by these algorithms. This algorithm also analyses the data sorting decisions by analysing if these decisions make the outcome worsens or improves the output (Baskin et al., 2016). Convolutional neural networks (CNNs), a type of ANN, contain neurons that are self-optimised through learning. (O'shea et al., 2021). They are classed as Deep Neural Networks, because CNNs have multiple layers (Alquraishi & Sorger, 2021). Recurrent neural network (RNN)

Table 1.1 Decision tree techniques

Decision tree technique	Method	Reference
Random decision forests	Construct multiple decision trees at the training stage. The final decision is the most common output.	(Ho, 1995)
Neighbour joining	Bottom-up method where outputs that are most similar are grouped together.	(Saitou & Nei, 1987)
Regression analysis	Nodes represent the mean of the results of the preceding nodes.	(Kamiński et al., 2018).
Binary Decision Tree	*Sequential decision process with features evaluated with one of two outcomes.*	(García Márquez et al., 2019)

remembers previous analyses both the inputs and resulting outcomes and then treats all future inputs and outputs as related (Dupond, 2019).

AI must be able to make decisions to perform its analysis and useful feature selection. The decision tools used are generally decision trees. These decision tools are named trees as the graphical representation of the decision-making process resembles a flowchart. The AI performs a test or analysis of each piece of data, and this gives rise to separate results. Each decision is represented as a node and each result represented as a branch. The final results then lead to a further analysis of each branch. This gives rise to the branched tree structure with some decisions (results proving to be dead ends). The final terminal nodes are known as classification or label (Kamiński et al., 2018). There are different types of trees as shown in Table 1.1 and Fig. 1.2.

AI has used decision trees to improve diagnosis. One study used lung cancer samples from the Lung Image Database Consortium (LIDC) dataset. This data was split 90% for training and 10% for testing. A labelled subset of the training set was used to train a CNN-based ransom decision tree. Once the CNN random decision tree was trained it was tested on the test data. This tree was able to accurately assign labels to the unlabeled data (Zheng et al., 2019). The origin of cancer tissue gas has been predicted based on miRNA profiling using AI based on two types of decision tree. Firstly, with neighbour joining methods and secondly with binary decision tree analyses. The neighbour joining method with an accuracy of 93.9%. The prediction accuracy of the binary decision tree method was 84.8% (Park et al., 2021).

Guidelines have been established in order to assist in the validation of the analysis provided by AI. These guidelines are known as the critical assessment of genome interpretation (CAGI) and were formulated using variants that were experimentally validated to cause disease and assessing if those predictions obeying the guidelines match these validated results (Andreoletti et al., 2019). The fifth edition of CAGI created in 2021 consists of 14 questions or criteria known as challenges (Andreoletti et al., 2019).
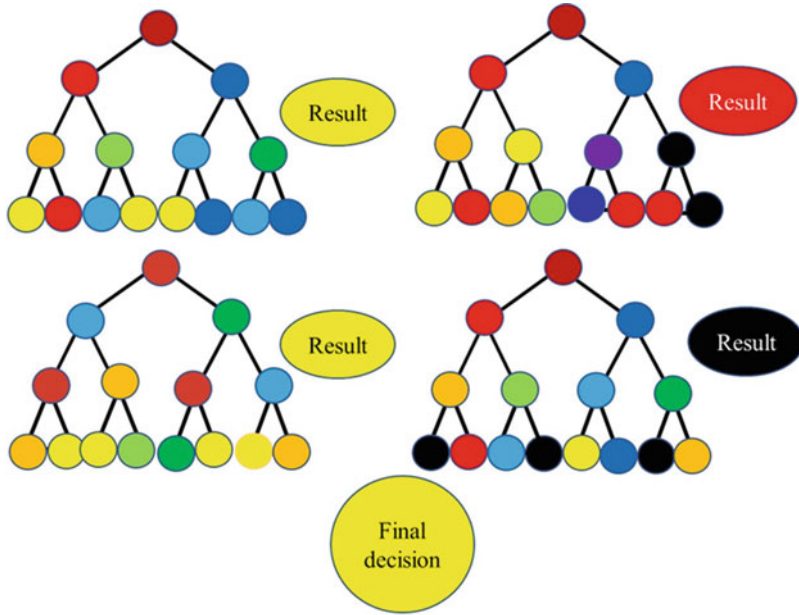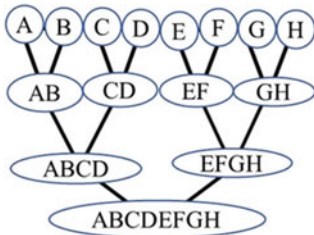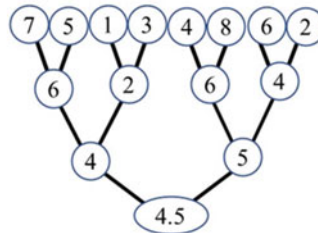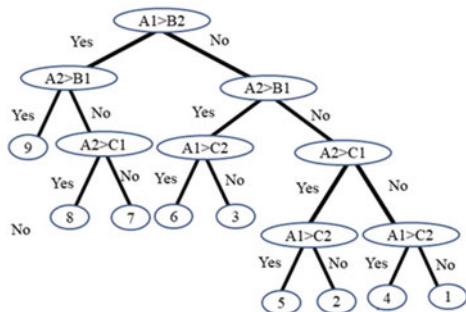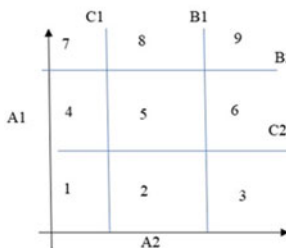
Fig. 1.2 Depictions of common decision tree methods (**a**) Random Forest trees use multiple trees and then select the most common outcome (**b**) Neighbour joining tree group nodes by similarity and select between these similar nodes (**c**) Regression trees the nodes are the mean of the previous nodes (**d**) Binary trees with sequential decisions based on one of two possible outcomes

1.3 Biomarker Discovery and Application

An ideal strategy to improve the screening, diagnosis, classification, staging and treatment of various cancers, is the identification of various molecules or molecular patterns or profiles that can serve as biomarkers. These biomarkers can be genomic mutations, transcripts, non-coding RNAs, proteins, metabolites, or even epigenetic markers. When patients present with symptoms indicating that they may have cancer, the standard procedure is a physical examination and radiographic imaging, this may be followed by biopsy examination. Many cancers screening procedures require invasive or expensive procedures. Biomarkers are normally classified as prognostic or predictive. Prognostic biomarkers are used to categorise patients by their risk of developing disease (screening), diagnosis of the disease, risk of disease progression, severity of disease and risk of death from the disease (Echle et al., 2021). Predictive biomarkers can be used to select a targeted treatment. These predictive biomarkers can also be used for drug discovery or in clinical trials for new treatments (Echle et al., 2021). The discovery of these biomarkers relies on the use of large omics datasets and the identification of patterns of the presence or absence of molecules in these large datasets that can be associated with disease. AI is a vital tool in this discovery process as it allows these large datasets to be rapidly and accurately analysed and associations with diseases to be identified. This is made possible through the use of machine and deep learning algorithms. Indeed, DL-based image analysis has broad applications in multiple fields of modern medicine that involve image data: in radiology, DL performs (Echle et al., 2021).

Liquid biopsies involve the identification of biomarkers in various body fluids. This can be blood, urine, saliva, or even cerebral spinal fluid. This is a more ideal diagnostic or prognostic technique than normal biopsies as they are less invasive and traumatising to a patient. This is also an important consideration for precision medicine as these samples can be obtained and analysed rapidly to give a current view of the patients' health and status (Kaur et al., 2017). These biomarkers can be transcripts, genomic markers in the form of DNA, proteins, or metabolites. In the case of RNA and DNA the transcripts would appear in biological fluids in the form of circulating cell free nucleic acids (ccfNAs). These ccfNAs have already been used as biomarkers in cancer diagnosis, prognosis, and monitoring (reviewed in (Pös et al., 2018)). It has also been established that these ccfNA appear in higher amounts in disorders such as cancer (Pös et al., 2018).

These nucleic acids can be in the form of cell free DNA which is fragmented DNA usually no longer than 450 bp in size. This DNA can be either of genomic or mitochondrial origin (Thierry et al., 2016). Circulating cell free RNAs include mRNA transcripts, non-coding RNAs, such as microRNA (miRNA), long non-coding RNA (lncRNA) and circular RNAs, transfer RNAs and ribosomal RNAs (reviewed in (Pös et al., 2018)). These nucleic acids are normally released into the body fluids as a result of cell death or in the case of many of the RNA molecules through active secretion (Vita et al., 2022).

1.4   Multi-omics Data

High throughput techniques like NGS allow for in-depth analysis of the mutational landscapes, gene expression patterns and epigenetic modifications for a large number of samples. Integration of "multi-omics" (genomics, epi-genomics, transcriptomics, proteomics, and metabolomics), and "non-omics" (medical/mass-spectrometry imaging, patient clinical history, treatments, and disease endemicity) data could help overcome the challenges in the accurate detection, characterisation, and monitoring of cancers. The complex analysis, annotation and combination of various omics data is sometimes only possible following data simplification. When these simplification processes are performed it is important to note that it may lead to the loss of information. The complexity of data is normally measured by the number of dimensions (variables) it has (Pezoulas et al., 2021). This reduction allows for increased ease and speed of analysis as well as a reduction in the space needed to store the data (Meng et al., 2016).

1.4.1   Genomics

The generation of large genomic datasets is due to advances in next-generation sequencer (NGS) (Paolillo et al., 2016)) and in silico computational algorithms. Whole genome sequencing allows for the analysis of all genomic alterations in cancer. It gives information regarding the number and identity of driver mutations and allows the mutational signature of the tumour to be identified. WGS has led to multiple sequencing projects and the establishment of databases containing the DNA sequence profiles of many cancers. These databases are listed in Table 1.2. To be truly useful genomic data must be integrated with clinical data, patient demographics, survival data, treatment status (Robinson et al., 2017). This is needed to link genomic events to specific cancers prognoses, and treatment responses (Robinson et al., 2017). AI has immense potential to contribute significantly at every stage of cancer management ranging from reliable early detection, stratification, determination of infiltrative tumour margins during surgical treatment, response to drugs/therapy, tracking tumour evolution and potential acquired resistance to treatments over time, prediction of tumour aggressiveness, metastasis pattern and recurrence (Bi et al., 2019).

1.4.2   Transcriptomics

Transcriptome includes the transcribed mRNAs, the alternately spliced isoforms of those mRNAs as well as non-coding RNAs such as miRNA. Any study looking at all these transcripts will aim to identify all the transcripts involved in metabolic

Table 1.2 DNA sequence databases and their applications

Database	Application	Reference
The Cancer Genome Atlas (TCGA)	Understand the molecular basis of cancer through the application of genomics.	(Wang, Jensen, & Zenklusen, 2016)
International Cancer Genome Consortium (ICGC)	Voluntary collaborative forum.	(Zhang et al., 2019)
CatLog of Somatic Mutations in Cancer (COSMIC)	Catalogue of somatic mutations in human cancer showing the impact of these mutations.	(Forbes et al., 2015)
The NCI's Genomic Data Commons (GDC)	A unified repository for cancer knowledge enabling data sharing across cancer genomic studies in support of precision medicine.	(Gao et al., 2013)
cBioPortal	Provides visualisation, analysis and download of large-scale cancer genomics data sets.	(Gao et al., 2013)
Methyl-Cancer	Database for human DNA Methylation and Cancer	(He et al., 2008)
UCSC Cancer Genomics Browser	Displays whole-*genome* and pathway-oriented views of genome-wide experimental measurements for individual and sets of samples.	(Goldman et al., 2013)
Moonshot project	Aims to address inequalities in access to cancer screening in the USA.	(Hsu et al., 2017)

processes and how they interact to result in gene expression. Studies that only examine specific sets of transcripts, i.e., mRNAs or miRNAs will provide answers to more specific questions. The result of epigenetic changes that occur in cancer can and have been studied by examining the transcriptome of cancers where these epigenetic changes have occurred. These studies have been undertaken in breast cancer (Robinson et al., 2015), prostate (Varambally et al., 2002) (Bhasin et al., 2015), head and neck squamous cell carcinoma (HNSCC) (Kelley et al., 2017).

1.4.3 Proteomics

Proteomic profiles reveal the actual cellular response to the conditions a cell is faced with. The change in protein expression also provides information regarding processes that affect protein modification, transport, and stability. Datasets of protein expression profiles are created using mass spectrometry and have been used to profile protein expression changes in response to therapy, monitor drug toxicity, and for diagnosis using specific biomarkers. These biomarker profiles, which are identified through protein expression signatures can also be to monitor disease progression, establish metastatic risk, do treatment follow-up to check for recurrence and stratify patients according to subtype (Keyl et al., 2022). Once again, these large data sets require AI to interpret them accurately, reliably, rapidly and consistently.

Many AI algorithms have been used to infer protein–protein interaction networks from proteomic datasets (Keyl et al., 2022). Another significant role for AI in proteomics is predicting docking capabilities between drugs and their target compounds.

AI can also be used to combine and integrate proteomic and genomic data to identify DNA mutations related to protein signalling. These genetic changes can then be said to be genetic drives of cancer. This has been performed in breast cancer, where the identification of signalling pathways specifically altered in different breast cancer subtypes was achieved. It also identified SKP1 and CETN3 as two new markers for basal-like breast cancer (Mertins et al., 2016). Proteomic and transcriptomic data can be integrated to identify changes in the splicing of mRNA and the generation of different protein isoforms that may be characteristic of different cancers (Liu et al., 2017). Proteomic data can show a much stronger association to the clinical characteristics of a patient, and this is reflected by the close association of integrated proteomics data with the clinical outcomes, for example MS analysis integrated with histopathological diagnosis (Huber et al., 2014). This can be done with very small amounts of extracted proteins, for example a study was performed where very small amounts of protein were analysed using LC-MS which led to deep coverage of entire proteomes of specific cell types (Kulak et al., 2017). A recent development has been the use of single-cell proteomics which has gained importance since it is able to give insights into cancer heterogeneity and the metastatic ability of single cells compared to colonies. It is also able to provide information concerning rare/mutated cells (Doerr, 2019). This has been successfully used to grade and rank acute myeloid leukaemia hierarchy (Schoof et al., 2021b).

1.4.4 Metabolomics

Metabolomics is the analysis of small molecules, such as amino acids, lipids, nucleotides, carbohydrates and organic acids, which are produced because of primary or secondary metabolic processes. The populations of these molecules changes during, growth, in response to stress and consequently during the development and progression of cancer (Bertini et al., 2009) (Lin et al., 2011) (Veselkov et al., 2011). Therefore, metabolomics can be used as an indicator of the molecular mechanisms underlying tumorigenesis.

It can also be used to monitor disease progression, the response of the tumour to drugs and other treatments. As with proteomics, the profiling of metabolites relies on mass spectrometry but with the additional use of nuclear magnetic resonance (NMR) spectroscopy (Merz & Serkova, 2009). Traditionally the sample had to be separated or fractionated to achieve the best results, but separation-free MS techniques have been developed which reduce the volume of sample required and reduce variation in the data generated through the analysis. These include direct infusion-MS, MALDI-MS, mass spectrometry imaging (MSI), and direct analysis in real-time mass spectrometry (Dettmer et al., 2007). The Global Natural Product Social Molecular

Networking (GNPS) is a *small-molecule mass spectrometry networking hub. Researchers can deposit their own MS data for small molecules and this repository is available for other uses to search and use*. GNPS has been shown to be very useful for cataloguing and organising MS/MS data using AI in the form of correlation and visualisation approaches. These can be used to identify spectra from related molecules (Wang, Carver, et al., 2016). Techniques such as principal component analysis or hierarchical clustering can be used in conjunction with ML to data mine these repositories to enhance the identification of spectra (Bertini et al., 2009) (Duan et al., 2005). These techniques have been used to identify metabolic biomarkers for multiple cancers including colorectal (Yamazaki, 2015), pancreatic (Zhang et al., 2012), lung (Zhuang et al., 2016), breast (Li et al., 2020), gastric (Ikeda et al., 2012), ovarian (Zhang et al., 2013) and prostate (Kelly et al., 2016).

1.4.5 Microbiomics

It has been estimated that the microbiota of the average human contains 40 trillion microbial cells (Sender et al., 2016). This microbiota is now known to play a role in the development and progression of cancer, especially through interactions with the nervous system and what is known as the gut–brain axis (reviewed in (Hull et al., 2021)). The profiling of all the microbial genes, metabolites, proteins and transcripts within a single patient is known as the patient's microbiome (Sepich-Poore et al., 2021). This can partly be explained by the interaction between the microbiome and the immune system as this may favour the development of cancer (Mangani et al., 2017). Microbiomes have been so closely associated with cancer, that it is now known that specific populations of microorganisms and microbial metabolites are associated with specific cancers. Therefore, different microbial signatures can be used as biomarkers to diagnose or monitor cancer, and affect the safety, tolerability and efficacy of specific treatments. Microbiomics are studies using the same high throughput techniques such as NGS and mass spectrometry. Once again this gives rise to large databases, which require the use of AI and machine or deep learning to analyse and interpret this data. Any attempt to integrate this microbiomic data with other "omics" data would require the use of AI (reviewed in (Cammarota et al., 2020)). AI can also be used to identify and evaluate microbiome community interactions with other microbes or the host. This is done using Network analysis and is useful for the identification of changes in these interactions may be caused by microbial community structure, environmental factors, metabolites, clinical. These networks can be constructed based on similarity or correlation coefficients between pairwise variables. Extended relationships can then be inferred based on these pairwise interactions. This is done using algorithms such as SparCC (Sparse Correlations for Compositional data) (Friedman & Alm, 2012) and Compositionally Robust Inference of Microbial Ecological Networks) (Faust et al., 2012).

1.5   Imaging

Medical imaging techniques such as Magnetic Resonance Imaging (MRI), CT scans, and Positron emission tomography (PET), are commonly used in the diagnosis of cancer. This is because these techniques are good at soft tissue contrast. This allows them to be good at locating tumours and monitoring tumour progression. They are also non-invasive and have a high resolution (Magadza & Viriri, 2021) (Menze et al., 2014). The aim of imaging cancer or suspected cancer tissue is known as tumour segmentation. This is the act of distinguishing between normal and cancerous tissue. This is a vital procedure for the use of imaging techniques in diagnosis and treatment planning, monitoring treatment response and disease progression (Bousselham et al., 2019). AI has been successfully used to automate the interpretation of medical imaging. It has been shown to be able to analyse stained sections of temper tissue and segment these images allowing for the identification and quantification of various parameters. These include the rate and amount of mitosis (Romo-Bucheli et al., 2017), the presence and abundance of mutations (Coudray et al., 2018), the differentiation between nuclei from benign cells versus those from cancer cells (Sirinukunwattana et al., 2016) (Xu et al., 2016), spatial localisation of proteins (Saltz et al., 2018). AI-based image analysis is more reproducible, objective and is quantitative compared to manual assessment. Convolutional neural networks (CNNs) are most commonly used for image analysis (Muhammad et al., 2020). There are two types of automated segmentation, generative and discriminating methods (Magadza & Viriri, 2021). Both methods use the same seven stages of analysis image acquisition, image preprocessing (deionising/enhancement/restoration), image segmentation/feature extraction and object recognition (Pan, 2007).

Image Segmentation techniques are all based on pixel-based selection to discern a Region of Interest (ROI). However, there are different methods that are used to achieve this, In the region-based method, a pixel in the ROI is selected as the reference or seed pixel. Neighbouring pixels are then compared to this pixel in order to establish if they are similar enough to be included (Punitha et al., 2018). In the edge-based method, the image is reduced to only its important structural characteristics. This decreases the image size. It also allows for the image's background to be separated from the object (Farag, 1992). The fuzzy theory-based method is an amalgamation of the region and the edge methods. (Basir et al., 2003). The partial differential equation (PDF) method calculates an energy of the image function. It then uses a partial differential equation (PDE) to describe the parametric curve evolution based on the energy of the image. It then uses this equation to find similar pixels (Sliž & Mikulka, 2016). In the threshold-based method, a grayscale binary image is created to reduce image complexity. This makes it easier to classify pixels (Bhargavi & Jyothi, 2014). Finally, the semantic segmentation network method classifies every individual pixel as either tumour or normal (Chen et al., 2017). When it comes to performing Whole Slide Image (WSI) segmentation some of these methods are more time and computing power consuming than others. The semantic method is the slowest and requires the most computing power (Guo et al., 2019).

1.5.1   Radiogenomics

Histopathological images have been integrated with genomics data in order to enhance feature selection based on cancer tissue architecture (López de Maturana et al., 2019). In a similar way, multi-omics data have been associated with features in medical images to develop predictive models using AI algorithms. This has been successfully performed for prostate cancer (Robinson et al., 2015), renal cell carcinoma (Schoof et al., 2021a), low-grade glioma (Brat et al., 2015), non-small cell lung cancer (Yu et al., 2016) and breast cancer (Yuan et al., 2012). This technique was initially given the name imaging genomics since it associated image features with genomic data. However, another term, radiomics or radiogenomics has been used to cover all the different omics data that can be associated with image features (Bodalal et al., 2019). Image features that can be associated with this omics data include structures, shapes, lines, points, colours or boundaries. It can even be extended to regions of the image associated with these features (Bi et al., 2019). In order to carry out a radiogenomics analysis, the AI must extract features identified on an image and link these features with phenotypes which is due to protein expression which can then be associated with genomic, transcriptomic and epigenomic or other omics changes (Rutman & Kuo, 2009). The appearance of these features on an image can then be an indication that these omics changes are present in the patient and the tumour. In the same way, these omics profiles can be used as indicators of for instance patient survival or disease progression, these associated image features can now be used to do the same (Berger & Mardis, 2018). AI is also necessary in radiogenomics as some of the features or changes in the cancer tissue may be so subtle that they may be missed by the human eye. Computer-assisted image analysis will accurately and consistently detect these changes based on what the algorithm has learned from previous data thanks to the application of machine and deep learning. These changes can then be associated accurately and without bias to any genomic, proteomic, transcriptomic, epigenomic, metabolomic or feature within the patient records. This is due to the analysis the AI can conduct on this data to extract unique features and then associate them with the unique image features. As previously stated, this integration would be too complex for a human being to complete accurately and timeously (Hussein et al., 2017). This end-to-end, automated data analysis or pipeline is able to compute and discriminate a vast number of features in both the image analysis and patient records or omics data to achieve the most accurate selection of features that are associated. And these models' ability to learn means that they are optimising their analytical ability and performance while integrating these data sets and looking for associations (Jansen et al., 2018).

Ai and radiogenomics have been shown to be able to predict the neoadjuvant therapy response in esophageal cancer using a convolutional neural network to analyse fluorodeoxyglucose positron emission tomography (18F-FDG PET) images. It was able to associate features from these images with transcriptomic data and make highly specific and accurate predictions (Ypsilantis et al., 2015). In another study, AI algorithms were used to identify image features within PET images in

breast cancer patients and associate these features with a genetic biomarker (Fujishima et al., 2017). Studies have also shown that features in images could be associated with tumour mutational burden, the average number of genetic mutations per megabase (Angus et al., 2019) and with the metastatic ability of the tumour (Trivizakis et al., 2019).

1.6 Drugs, AI and Precision Oncology

1.6.1 *Drug Discovery and Re-purposing*

The design or discovery of new drugs is a time-consuming and expensive undertaking with many potential compounds that have already had large amounts of money, $314 million to $2.8 billion, spent on them failing in the final stages. This means that all the time and money spent on them was essentially wasted (Waring et al., 2015). It is estimated that 90% of drugs fail to enter clinical trials for regulatory approval in (Fleming, 2018). AI can be used to remove those compounds most likely to fail from further development and prevent resources being wasted on them (Gawehn et al., 2016). This can be achieved using modelling to design better drugs by assessing a compound's binding abilities, identifying their binding partners that may be biologically significant and establishing if there are any toxic interactions they may have. Some of these modelling algorithms that have been developed and that are already in use include the quantitative structure-activity relationship (QSAR) models. These models still face problems since they need to learn from experimental data sets. If these datasets are small, it may decrease the accuracy of the model. If the data is not validated there may be errors that would lead to errors in the final model due to the algorithm learning from incorrect data (Roy & Pratim Roy, 2009) (Zhao et al., 2017). AI can also be used to search through chemical databases to identify compounds with a structure that may indicate their ability to bind to a specific target. The searching of these large libraries is known as high-throughput screening techniques (HTS) (Inglese et al., 2006) (Zhu et al., 2016).

AI can also be used to predict how a drug will behave with respect to its physicochemical properties, bioactivity and toxicity. Physiochemical properties can be predicted using AI-based tools such as using Quantitative Structure Property Relationship (QSPR) workflow. This algorithm was originally designed to predict the physiochemical properties of environmental toxins (Zang et al., 2017). Other algorithms have also been developed that are able to perform function such as predicting the solubility of a drug, these include undirected graph recursive neural networks and graph-based convolutional neural networks (CVNN) (Kumar et al., 2017). The efficacy of drugs can be predicted by establishing their affinity for their target molecule, and toxicity and side effects may be predicted by identifying any unintended interactions it may have. AI is able to accomplish these actions by calculating the binding affinities for the drug on a large number of molecules. It can do this by identifying any similar features or structures the drug has with similar

molecules or targets similar to the intended target (Öztürk et al., 2018). Screening for the most effective treatment for a specific patient is also possible using AI. One way this can be done is through the use of a digital twin.

1.6.2 Digital Twins

An important concept in the use of AI in medicine is the creation of a digital twin. This digital twin is the use of patient-specific data to create a virtual copy of the patient. An accurate digital twin requires accurate, detailed and up-to-date information about the patient (Batch et al., 2022). Deciding on the best treatment for an individual patient is one of the primary uses of the digital twin. This process is demonstrated in Fig. 1.3. As much data concerning the patient is gathered. This includes various omics data, patient records, medical imaging and imaging reports and any data concerning demographic or risk factors. AI then creates the digital twin. These twins are then duplicated, and each twin is given a virtual treatment. Using information regarding the molecular basis of these treatments, their side effects and case reports and studies of these treatments and AI algorithm can then run simulations for each individual treatment on the digital twin. The results can be used to select the best treatment option (Björnsson et al., 2019).

There are many ways these drugs can be tested in these simulations. One example is the use of protein–protein interaction (PPI) networks, constructed using a patient's proteomic or transcriptomic data as a map. Changes in protein expression caused by
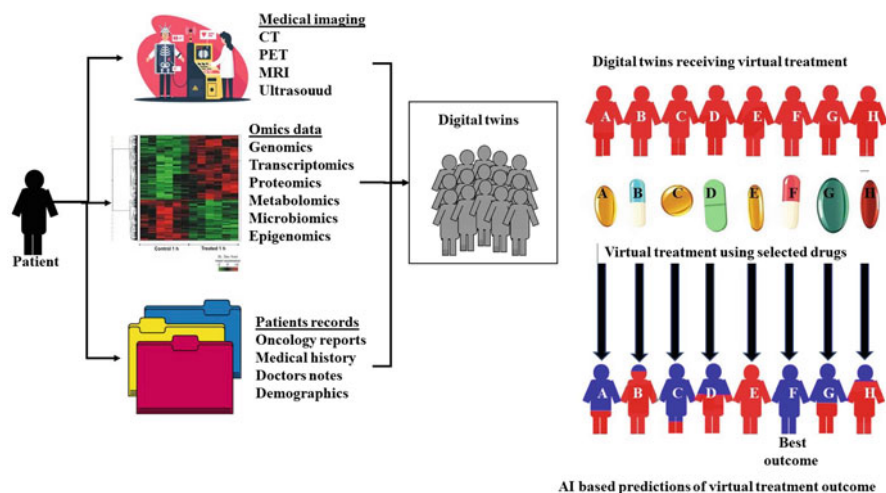

Fig. 1.3 The use of digital twins in drug discovery. Various types of data from a patient are used to create the most accurate digital twin possible. This twin is then duplicated and treated virtually with all available drugs. Artificial intelligence then calculates treatment outcomes based on drug molecular interactions and the molecular environment of the digital twin

a treatment can then be mapped to the patients PPI to identify changes in the pathways the drug could cause when used to treat the patient (Barabási et al., 2011) (Zhou et al., 2014). Another example could involve genetic changes detected in a patient. These alterations that lead to transcript and protein changes can be used to create a twin with the altered protein and protein expression patterns. A treatment targeting this protein can be used to treat the digital twin. The resulting effects on PPI and pathways can then be simulated in the twin.

1.7   Conclusion

The integration and analysis of data from various sources such as different "omics", medical images and medical imaging reports, electronic medical records, or hand-written doctor's notes, is only possible in a practical manner using AI and machine learning. The requirement for the use of AI has been necessitated due to the advancements in multidimensional "omics" technologies. The application of AI to biological data enables the understanding of complex biological systems. AI is already used in the automated extraction of information as well as the automated integration of health records. It is also currently used to organise, annotate and store data in big data storage systems such as cloud scaling. AI can outperform human clinicians and pathologists in all these tasks and it has enabled us to develop new techniques to study cancer, detect cancer at an early stage, more accurately predict patient outcomes. Decide on the correct treatment, monitor disease progression and treatment effectiveness, design new drugs and therapies, and stratify and classify tumours (Fig. 1.4).

This chapter has served as a brief introduction to the various topics that will be covered in the following chapters of this book. The initial chapters will examine the use of AI in the identification and application of novel biomarkers for precision oncology. This involves the use of these biomarkers in diagnosis, screening, monitoring drug resistance and in the choice of the most appropriate treatment regimen. They will also discuss the novel use of ccfNAs as biomarkers in precision medicine. The last of these initial chapters will discuss the use of digital pathology in accomplishing these tasks and how the new field of radiogenomics will allow image features to be associated with molecular signatures. The book will then discuss some of the less discussed "omics" that are studied to obtain data that can be used to identify biomarkers for use in precision oncology. These include epigenomics, metabolomics and microbiomics. The final chapters of the book will discuss the practical and clinical application of AI to precision oncology in detail. The first of these applications the book will discuss is the use of nanotechnology in AI-based precision oncology. It will then focus on the use of AI-based devices in cancer screening. This will be followed by a chapter describing the use of AI to design new drugs and then a chapter describing the application of AI to increase the efficiency of immunotherapy. For the final applications, the book will discuss the role AI can play in helping clinicians and oncologists choose the correct treatment

Fig. 1.4 A summary of the applications of AI in precision oncology

for an individual patient using various AI tools and techniques. The concluding chapter will summarise the topics covered and offer insights into the future of AI in precision oncology.

References

Alquraishi, M., & Sorger, P. K. (2021). Differentiable biology: Using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nature Methods, 18*, 1169–1180.

Andreoletti, G., Pal, L. R., Moult, J., & Brenner, S. E. (2019). Reports from the fifth edition of CAGI: The critical assessment of genome interpretation. *Human Mutation, 40*, 1197–1201.

Angus, L., Smid, M., Wilting, S. M., Van Riet, J., Van Hoeck, A., Nguyen, L., Nik-Zainal, S., Steenbruggen, T. G., Tjan-Heijnen, V. C. G., Labots, M., Van Riel, J., Bloemendal, H. J., Steeghs, N., Lolkema, M. P., Voest, E. E., Van de Werken, H. J. G., Jager, A., Cuppen, E., Sleijfer, S., & Martens, J. W. M. (2019). The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nature Genetics, 51*, 1450–1458.

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews. Genetics, 12*, 56–68.

Basir, O., Zhu, H., & Karray, F. (2003). Fuzzy based image segmentation. In M. Nachtegael (Ed.), *Fuzzy filters for image processing*. Springer.

Baskin, I. I., Winkler, D., & Tetko, I. V. (2016). A renaissance of neural networks in drug discovery. *Expert Opinion on Drug Discovery, 11*, 785–795.

Batch, K. E., Yue, J., Darcovich, A., Lupton, K., Liu, C. C., Woodlock, D. P., El Amine, M. A. K., Causa-Andrieu, P. I., Gazit, L., Nguyen, G. H., Zulkernine, F., Do, R. K. G., & Simpson, A. L. (2022). Developing a cancer digital twin: Supervised metastases detection from consecutive structured radiology reports. *Frontiers in Artificial Intelligence, 5*, 826402–826402.

Berger, M. F., & Mardis, E. R. (2018). The emerging clinical relevance of genomics in cancer medicine. *Nature Reviews. Clinical Oncology, 15*, 353–365.

Bertini, I., Calabrò, A., De Carli, V., Luchinat, C., Nepi, S., Porfirio, B., Renzi, D., Saccenti, E., & Tenori, L. (2009). The metabonomic signature of celiac disease. *Journal of Proteome Research, 8*, 170–177.

Bhargavi, K., & Jyothi, S. (2014). A survey on threshold based segmentation technique in image processing. *International Journal of Innovative Research & Development, 3*, 234–239.

Bhasin, J. M., Lee, B. H., Matkin, L., Taylor, M. G., Hu, B., Xu, Y., Magi-Galluzzi, C., Klein, E. A., & Ting, A. H. (2015). Methylome-wide sequencing detects DNA hypermethylation distinguishing indolent from aggressive prostate cancer. *Cell Reports, 13*, 2135–2146.

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F., Mak, R. H., Tamimi, R. M., Tempany, C. M., Swanton, C., Hoffmann, U., Schwartz, L. H., Gillies, R. J., Huang, R. Y., & Aerts, H. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: a Cancer Journal for Clinicians, 69*, 127–157.

Björnsson, B., Borrebaeck, C., Elander, N., Gasslander, T., Gawel, D. R., Gustafsson, M., Jörnsten, R., Lee, E. J., Li, X., Lilja, S., Martínez-Enguita, D., Matussek, A., Sandström, P., Schäfer, S., Stenmarker, M., Sun, X. F., Sysoev, O., Zhang, H., & Benson, M. (2019). Digital twins to personalize medicine. *Genome Medicine, 12*, 4.

Bodalal, Z., Trebeschi, S., Nguyen-Kim, T. D. L., Schats, W., & Beets-Tan, R. (2019). Radiogenomics: Bridging imaging and genomics. *Abdominal Radiology (New York), 44*, 1960–1984.

Bousselham, A., Bouattane, O., Youssfi, M., & Raihani, A. (2019). Towards reinforced brain tumor segmentation on MRI images based on temperature changes on pathologic area. *International Journal of Biomedical Imaging, 3*, 1758948.

Brat, D. J., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., Cooper, L. A., Rheinbay, E., Miller, C. R., Vitucci, M., Morozova, O., Robertson, A. G., Noushmehr, H., Laird, P. W., Cherniack, A. D., Akbani, R., Huse, J. T., Ciriello, G., Poisson, L. M., Barnholtz-Sloan, J. S., Berger, M. S., Brennan, C., Colen, R. R., Colman, H., Flanders, A. E., Giannini, C., Grifford, M., Iavarone, A., Jain, R., Joseph, I., Kim, J., Kasaian, K., Mikkelsen, T., Murray, B. A., O'Neill, B. P., Pachter, L., Parsons, D. W., Sougnez, C., Sulman, E. P., Vandenberg, S. R., Van Meir, E. G., Von Deimling, A., Zhang, H., Crain, D., Lau, K., Mallery, D., Morris, S., Paulauskis, J., Penny, R., Shelton, T., Sherman, M., Yena, P., Black, A., Bowen, J., Dicostanzo, K., Gastier-Foster, J., Leraas, K. M., Lichtenberg, T. M., Pierson, C. R., Ramirez, N. C., Taylor, C., Weaver, S., Wise, L., Zmuda, E., Davidsen, T., Demchok, J. A., Eley, G., Ferguson, M. L., Hutter, C. M., Mills Shaw, K. R., Ozenberger, B. A., Sheth, M., Sofia, H. J., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J. C., Ayala, B., Baboud, J., Chudamani, S., Jensen, M. A., Liu, J., Pihl, T., Raman, R., Wan, Y., Wu, Y., Ally, A., Auman, J. T., Balasundaram, M., Balu, S., Baylin, S. B., Beroukhim, R., Bootwalla, M. S., Bowlby, R., Bristow, C. A., Brooks, D., Butterfield, Y., Carlsen, R., Carter, S., Chin, L., Chu, A., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *The New England Journal of Medicine, 372*, 2481–2498.

Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., Claesson, M. J., & gasbarrini, A. & Tortora, G. (2020). Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nature Reviews Gastroenterology & Hepatology, 17*, 635–648.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*, 834–848.

Cheng, D., Knox, C., Young, N., Stothard, P., Damaraju, S., & Wishart, D. S. (2008). PolySearch: A web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. *Nucleic Acids Research, 36*, W399–W405.

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A. L., Razavian, N., & Tsirigos, A. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine, 24*, 1559–1567.

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews, 26*, 51–78.

Doerr, A. (2019). Single-cell proteomics. *Nature Methods, 16*, 20.

Duan, K. B., Rajapakse, J. C., Wang, H., & Azuaje, F. (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Transactions on Nanobioscience, 4*, 228–234.

Dupond, S. (2019). A thorough review on the current advance of neural network structures. *Annual Reviews in Control, 14*, 200–230.

Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., & Kather, J. N. (2021). Deep learning in cancer pathology: A new generation of clinical biomarkers. *British Journal of Cancer, 124*, 686–696.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., Diester, I., Brox, T., & Ronneberger, O. (2019). U-Net: Deep learning for cell counting, detection, and morphometry. *Nature Methods, 16*, 67–70.

Farag, A. A. (1992). Edge-based image segmentation. *Remote Sensing Reviews, 6*, 95–121.

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology, 8*, e1002606.

Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature, 557*, S55–s57.

Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., Mcdermott, U., & Campbell, P. J. (2015). COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research, 43*, D805–D811.

Friedman, J., & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Computational Biology, 8*, e1002687.

Fujishima, H., Fumoto, S., Shibata, T., Nishiki, K., Tsukamoto, Y., Etoh, T., Moriyama, M., Shiraishi, N., & Inomata, M. (2017). A 17-molecule set as a predictor of complete response to neoadjuvant chemotherapy with docetaxel, cisplatin, and 5-fluorouracil in esophageal cancer. *PLoS One, 12*, e0188098.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling, 6*, 11.

García Márquez, F. P., Segovia Ramírez, I., & Pliego Marugán, A. (2019). Decision making using logical decision tree and binary decision diagrams: A real case study of wind turbine manufacturing. *Energies, 12*, 1753.

Gawehn, E., Hiss, J. A., & Schneider, G. (2016). Deep learning in drug discovery. *Molecular Informatics, 35*, 3–14.

Goldman, M., Craft, B., Swatloski, T., Ellrott, K., Cline, M., Diekhans, M., Ma, S., Wilks, C., Stuart, J., Haussler, D., & Zhu, J. (2013). The UCSC cancer genomics browser: Update 2013. *Nucleic Acids Research, 41*, D949–D954.

Govindarajan, R., Duraiyan, J., Kaliyappan, K., & Palanisamy, M. (2012). Microarray and its applications. *Journal of Pharmacy & Bioallied Sciences, 4*, S310–S312.

Guo, Z., Liu, H., Ni, H., Wang, X., Su, M., Guo, W., Wang, K., Jiang, T., & Qian, Y. J. S. R. (2019). A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Scientific Reports, 9*, 1–10.

Hakenberg, J., Voronov, D., Nguyên, V. H., Liang, S., Anwar, S., Lumpkin, B., Leaman, R., Tari, L., & Baral, C. (2012). A SNPshot of PubMed to associate genetic variants with drugs, diseases, and adverse reactions. *Journal of Biomedical Informatics, 45*, 842–850.

He, X., Chang, S., Zhang, J., Zhao, Q., Xiang, H., Kusonmano, K., Yang, L., Sun, Z. S., Yang, H., & Wang, J. (2008). MethyCancer: The database of human DNA methylation and cancer. *Nucleic Acids Research, 36*, D836–D841.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (pp. 278–282). IEEE.

Hsu, E. R., Klemm, J. D., Kerlavage, A. R., Kusnezov, D., & Kibbe, W. A. (2017). Cancer Moonshot data and technology team: enabling a national learning healthcare system for cancer to unleash the power of data. *Clinical Pharmacology and Therapeutics, 101*, 613–615.

Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters, 471*, 61–71.

Huber, K., Feuchtinger, A., Borgmann, D. M., Li, Z., Aichler, M., Hauck, S. M., Zitzelsberger, H., Schwaiger, M., Keller, U., & Walch, A. (2014). Novel approach of MALDI drug imaging, immunohistochemistry, and digital image analysis for drug distribution studies in tissues. *Analytical Chemistry, 86*, 10568–10575.

Hull, R., Lolas, G., Makrogkikas, S., Jensen, L. D., Syrigos, K. N., Evangelou, G., Padayachy, L., Egbor, C., Mehrotra, R., Makhafola, T. J., Oyomno, M., & Dlamini, Z. (2021). Microbiomics in collusion with the nervous system in carcinogenesis: Diagnosis, pathogenesis and treatment. *Microorganisms, 9*, 2129.

Hussein, S., Green, A., Watane, A., Reiter, D., Chen, X., Papadakis, G. Z., Wood, B., Cypess, A., Osman, M., & Bagci, U. (2017). Automatic segmentation and quantification of white and Brown adipose tissues from PET/CT scans. *IEEE Transactions on Medical Imaging, 36*, 734–744.

Ibrahim, A., Gamble, P., Jaroensri, R., Abdelsamea, M. M., Mermel, C. H., Chen, P. C., & Rakha, E. A. (2020). Artificial intelligence in digital breast pathology: Techniques and applications. *Breast, 49*, 267–273.

Ikeda, A., Nishiumi, S., Shinohara, M., Yoshie, T., Hatano, N., Okuno, T., Bamba, T., Fukusaki, E., Takenawa, T., Azuma, T., & Yoshida, M. (2012). Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. *Biomedical Chromatography, 26*, 548–558.

Inglese, J., Auld, D. S., Jadhav, A., Johnson, R. L., Simeonov, A., Yasgar, A., Zheng, W., & Austin, C. P. (2006). Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 11473–11478.

Jansen, R. W., Van Amstel, P., Martens, R. M., Kooi, I. E., Wesseling, P., De Langen, A. J., Der Houven, M.-V., Van Oordt, C. W., Jansen, B. H. E., Moll, A. C., Dorsman, J. C., Castelijns, J. A., De Graaf, P., & De Jong, M. C. (2018). Non-invasive tumor genotyping using radiogenomic biomarkers, a systematic review and oncology-wide pathway analysis. *Oncotarget, 9*, 20134–20155.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*, 230–243.

Joshi, S., Vibhute, G., Ayachit, A., & Ayachit, G. (2021). Big data and artificial intelligence - tools to be future ready? *Indian Journal of Ophthalmology, 69*, 1652–1653.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*, 237–285.

Kamiński, B., Jakubczyk, M., & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research, 26*, 135–159.

Kaur, J., Rahat, B., Thakur, S., Verma, M., & Barh, D. (2017). *Progress and challenges in precision medicine*. Academic Press.

Kehl, K. L., Xu, W., Gusev, A., Bakouny, Z., Choueiri, T. K., Riaz, I. B., Elmarakeby, H., Van Allen, E. M., & Schrag, D. (2021). Artificial intelligence-aided clinical annotation of a large multi-cancer genomic dataset. *Nature Communications, 12*, 7304.

Kelley, D. Z., Flam, E. L., Izumchenko, E., Danilova, L. V., Wulf, H. A., Guo, T., Singman, D. A., Afsari, B., Skaist, A. M., Considine, M., Welch, J. A., Stavrovskaya, E., Bishop, J. A., Westra, W. H., Khan, Z., Koch, W. M., Sidransky, D., Wheelan, S. J., Califano, J. A., Favorov, A. V., Fertig, E. J., & Gaykalova, D. A. (2017). Integrated analysis of whole-genome ChIP-Seq and RNA-Seq data of primary head and neck tumor samples associates HPV integration sites with open chromatin marks. *Cancer Research, 77*, 6538–6550.

Kelly, R. S., Vander Heiden, M. G., Giovannucci, E., & Mucci, L. A. (2016). Metabolomic biomarkers of prostate cancer: Prediction, diagnosis, progression, prognosis, and recurrence. *Cancer Epidemiology, Biomarkers & Prevention, 25*, 887–906.

Keyl, P., Bockmayr, M., Heim, D., Dernbach, G., Montavon, G., Müller, K. R., & Klauschen, F. (2022). Patient-level proteomic network prediction by explainable artificial intelligence. *NPJ Precision Oncology, 6*, 35.

Kulak, N. A., Geyer, P. E., & Mann, M. (2017). Loss-less Nano-fractionator for high sensitivity, high coverage proteomics. *Molecular & Cellular Proteomics, 16*, 694–705.

Kumar, R., Sharma, A., Siddiqui, M. H., & Tiwari, R. K. (2017). Prediction of human intestinal absorption of compounds using artificial intelligence techniques. *Current Drug Discovery Technologies, 14*, 244–254.

Kuwahara, T., Hara, K., Mizuno, N., Haba, S., Okuno, N., Koda, H., Miyano, A., & Fumihara, D. (2021). Current status of artificial intelligence analysis for endoscopic ultrasonography. *Digestive Endoscopy, 33*, 298–305.

Li, L., Zheng, X., Zhou, Q., Villanueva, N., Nian, W., Liu, X., & Huan, T. (2020). Metabolomics-based discovery of molecular signatures for triple negative breast cancer in Asian female population. *Scientific Reports, 10*, 370.

Lin, X., Wang, Q., Yin, P., Tang, L., Tan, Y., Li, H., Yan, K., & Xu, G. J. M. (2011). A method for handling metabonomics data from liquid chromatography/mass spectrometry: Combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection. *Metabolomics, 7*, 549–558.

Liu, Y., Gonzàlez-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., Venkitaraman, A. R., & Wickramasinghe, V. O. (2017). Impact of alternative splicing on the human proteome. *Cell Reports, 20*, 1229–1241.

López de Maturana, E., Alonso, L., Alarcón, P., Martín-Antoniano, I. A., Pineda, S., Piorno, L., Calle, M. L., & Malats, N. (2019). Challenges in the integration of omics and non-omics data. *Genes (Basel), 10*, 238.

Magadza, T., & Viriri, S. (2021). Deep learning for brain tumor segmentation: a survey of state-of-the-art. *Journal of Imaging, 7*, 19.

Mangani, D., Weller, M., & Roth, P. (2017). The network of immunosuppressive pathways in glioblastoma. *Biochemical Pharmacology, 130*, 1–9.

Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., & Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics, 17*, 628–641.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., & Wiest, R. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging, 34*, 1993–2024.

Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., Huang, K. L., Lin, C., Mclellan, M. D., Yan, P., Davies, S. R., Townsend, R. R., Skates, S. J., Wang, J., Zhang, B., Kinsinger, C. R., Mesri, M.,

Rodriguez, H., Ding, L., Paulovich, A. G., Fenyö, D., Ellis, M. J., & Carr, S. A. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature, 534*, 55–62.

Merz, A. L., & Serkova, N. J. (2009). Use of nuclear magnetic resonance-based metabolomics in detecting drug resistance in cancer. *Biomarkers in Medicine, 3*, 289–306.

Morin, O., Vallières, M., Braunstein, S., Ginart, J. B., Upadhaya, T., Woodruff, H. C., Zwanenburg, A., Chatterjee, A., Villanueva-Meyer, J. E., Valdes, G., Chen, W., Hong, J. C., Yom, S. S., Solberg, T. D., Löck, S., Seuntjens, J., Park, C., & Lambin, P. (2021). An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. *Nature Cancer, 2*, 709–722.

Muhammad, K., Khan, S., Del Ser, J., & De Albuquerque, V. H. C. (2020). Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey. *IEEE Transactions on Neural Networks and Learning Systems, 32*, 507–522.

O'shea, R. J., Sharkey, A. R., Cook, G. J. R., & Goh, V. (2021). Systematic review of research design and reporting of imaging studies applying convolutional neural networks for radiological cancer diagnosis. *European Radiology, 31*, 7969–7983.

Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). DeepDTA: Deep drug-target binding affinity prediction. *Bioinformatics, 34*, i821–i829.

Pan, Y. (2007). *Image segmentation using PDE, variational, morphological and probabilistic methods*. University of Tennessee.

Paolillo, C., Londin, E., & Fortina, P. (2016). Next generation sequencing in cancer: Opportunities and challenges for precision cancer medicine. *Scandinavian Journal of Clinical and Laboratory Investigation. Supplementum, 245*, S84–S91.

Park, J. W., Jeong, J. M., Cho, K. S., Cho, S. Y., Cheon, J. H., Choi, D. H., Park, S. J., & Kim, H. K. (2021). MiR-30a and miR-200c differentiate cholangiocarcinomas from gastrointestinal cancer liver metastases. *PLoS One, 16*, e0250083.

Pezoulas, V. C., Hazapis, O., Lagopati, N., Exarchos, T. P., Goules, A. V., Tzioufas, A. G., Fotiadis, D. I., Stratis, I. G., Yannacopoulos, A. N., & Gorgoulis, V. G. (2021). Machine learning approaches on high throughput NGS data to unveil mechanisms of function in biology and disease. *Cancer Genomics Proteomics, 18*, 605–626.

Pös, O., Biró, O., Szemes, T., & Nagy, B. (2018). Circulating cell-free nucleic acids: Characteristics and applications. *European Journal of Human Genetics, 26*, 937–945.

Punitha, S., Amuthan, A., & Joseph, K. S. (2018). Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal, 3*, 348–358.

Robinson, D., Van Allen, E. M., Wu, Y. M., Schultz, N., Lonigro, R. J., Mosquera, J. M., Montgomery, B., Taplin, M. E., Pritchard, C. C., Attard, G., Beltran, H., Abida, W., Bradley, R. K., Vinson, J., Cao, X., Vats, P., Kunju, L. P., Hussain, M., Feng, F. Y., Tomlins, S. A., Cooney, K. A., Smith, D. C., Brennan, C., Siddiqui, J., Mehra, R., Chen, Y., Rathkopf, D. E., Morris, M. J., Solomon, S. B., Durack, J. C., Reuter, V. E., Gopalan, A., Gao, J., Loda, M., Lis, R. T., Bowden, M., Balk, S. P., Gaviola, G., Sougnez, C., Gupta, M., Yu, E. Y., Mostaghel, E. A., Cheng, H. H., Mulcahy, H., True, L. D., Plymate, S. R., Dvinge, H., Ferraldeschi, R., Flohr, P., Miranda, S., Zafeiriou, Z., Tunariu, N., Mateo, J., Perez-Lopez, R., Demichelis, F., Robinson, B. D., Schiffman, M., Nanus, D. M., Tagawa, S. T., Sigaras, A., Eng, K. W., Elemento, O., Sboner, A., Heath, E. I., Scher, H. I., Pienta, K. J., Kantoff, P., De Bono, J. S., Rubin, M. A., Nelson, P. S., Garraway, L. A., Sawyers, C. L., & Chinnaiyan, A. M. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell, 161*, 1215–1228.

Robinson, D. R., Wu, Y. M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V., Schuetze, S., Alva, A., Siddiqui, J., Chugh, R., Worden, F., Zalupski, M. M., Innis, J., Mody, R. J., Tomlins, S. A., Lucas, D., Baker, L. H., Ramnath, N., Schott, A. F., Hayes, D. F., Vijai, J., Offit, K., Stoffel, E. M., Roberts, J. S., Smith, D. C., Kunju, L. P., Talpaz, M., Cieślik, M., & Chinnaiyan, A. M. (2017). Integrative clinical genomics of metastatic cancer. *Nature, 548*, 297–303.

Romo-Bucheli, D., Janowczyk, A., Gilmore, H., Romero, E., & Madabhushi, A. (2017). A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytometry. Part A, 91*, 566–573.

Roy, K., & Pratim Roy, P. (2009). Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *European Journal of Medicinal Chemistry, 44*, 2913–2922.

Rutman, A. M., & Kuo, M. D. (2009). Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging. *European Journal of Radiology, 70*, 232–241.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*, 406–425.

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K. R., Zhao, T., Batiste, R., Van Arnam, J., Shmulevich, I., Rao, A. U. K., Lazar, A. J., Sharma, A., & Thorsson, V. (2018). Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Reports, 23*, 181–193.e7.

Santus, E., Li, C., Yala, A., Peck, D., Soomro, R., Faridi, N., Mamshad, I., Tang, R., Lanahan, C. R., Barzilay, R., & Hughes, K. (2019). Do neural information extraction algorithms generalize across institutions? *JCO Clinical Cancer Informatics, 3*, 1–8.

Schoof, E. M., Furtwängler, B., Üresin, N., Rapin, N., Savickas, S., Gentil, C., Lechman, E., Auf Dem Keller, U., Dick, J. E., & Porse, B. T. (2021a). Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nature Communications, 12*, 745679.

Schoof, E. M., Furtwängler, B., Üresin, N., Rapin, N., Savickas, S., Gentil, C., Lechman, E., Auf Dem Keller, U., Dick, J. E., & Porse, B. T. (2021b). Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *bioRxiv, 12*, 745679.

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology, 14*, e1002533.

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., & Knight, R. (2021). The microbiome and human cancer. *Science, 371*, eabc4552.

Sirinukunwattana, K., Ahmed Raza, S. E., Yee-Wah, T., Snead, D. R., Cree, I. A., & Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging, 35*, 1196–1206.

Sliž, J. & Mikulka, J. (2016). Advanced image segmentation methods using partial differential equations: A concise comparison. In *2016 Progress in Electromagnetic Research Symposium (PIERS)* (pp. 1809–1812). IEEE.

Thierry, A. R., El Messaoudi, S., Gahan, P. B., Anker, P., & Stroun, M. (2016). Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Reviews, 35*, 347–376.

Trivizakis, E., Manikis, G. C., Nikiforaki, K., Drevelegas, K., Constantinides, M., Drevelegas, A., & Marias, K. (2019). Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation. *IEEE Journal of Biomedical and Health Informatics, 23*, 923–930.

Varambally, S., Dhanasekaran, S. M., Zhou, M., Barrette, T. R., Kumar-Sinha, C., Sanda, M. G., Ghosh, D., Pienta, K. J., Sewalt, R. G., Otte, A. P., Rubin, M. A., & Chinnaiyan, A. M. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature, 419*, 624–629.

Veselkov, K. A., Vingara, L. K., Masson, P., Robinette, S. L., Want, E., Li, J. V., Barton, R. H., Boursier-Neyret, C., Walther, B., Ebbels, T. M., Pelczer, I., Holmes, E., Lindon, J. C., & Nicholson, J. K. (2011). Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Analytical Chemistry, 83*, 5864–5872.

Vita, G. M., De Simone, G., De Marinis, E., Nervi, C., Ascenzi, P., & Di Masi, A. (2022). *Serum albumin and nucleic acids biodistribution: From molecular aspects to biotechnological applications*. IUBMB Life.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., Mclean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., Mcphail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nature Biotechnology, 34*, 828–837.

Wang, Z., Jensen, M. A., & Zenklusen, J. C. (2016). A practical guide to the cancer genome atlas (TCGA). *Methods in Molecular Biology, 1418*, 111–141.

Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Pairaudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., Wallace, O., & Weir, A. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature Reviews. Drug Discovery, 14*, 475–486.

Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., & Madabhushi, A. (2016). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Transactions on Medical Imaging, 35*, 119–130.

Yamazaki, Y. (2015). Metabolome analysis of human serum: Implications for early detection of colorectal cancer. *Rinsho Byori, 63*, 328–335.

Ypsilantis, P. P., Siddique, M., Sohn, H. M., Davies, A., Cook, G., Goh, V., & Montana, G. (2015). Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. *PLoS One, 10*, e0137036.

Yu, K. H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L., & Snyder, M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications, 7*, 12474.

Yuan, Y., Failmezger, H., Rueda, O. M., Ali, H. R., Gräf, S., Chin, S. F., Schwarz, R. F., Curtis, C., Dunning, M. J., Bardwell, H., Johnson, N., Doyle, S., Turashvili, G., Provenzano, E., Aparicio, S., Caldas, C., & Markowetz, F. (2012). Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Science Translational Medicine, 4*, 157ra143.

Zang, Q., Mansouri, K., Williams, A. J., Judson, R. S., Allen, D. G., Casey, W. M., & Kleinstreuer, N. C. (2017). In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. *Journal of Chemical Information and Modeling, 57*, 36–49.

Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L. D., & Ferretti, V. (2019). The international cancer genome consortium data portal. *Nature Biotechnology, 37*, 367–369.

Zhang, L., Jin, H., Guo, X., Yang, Z., Zhao, L., Tang, S., Mo, P., Wu, K., Nie, Y., Pan, Y., & Fan, D. (2012). Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by (1)H nuclear magnetic resonance-based metabonomic profiles. *Clinical Biochemistry, 45*, 1064–1069.

Zhang, T., Wu, X., Ke, C., Yin, M., Li, Z., Fan, L., Zhang, W., Zhang, H., Zhao, F., Zhou, X., Lou, G., & Li, K. (2013). Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. *Journal of Proteome Research, 12*, 505–512.

Zhao, L., Wang, W., Sedykh, A., & ZHU, H. (2017). Experimental errors in QSAR Modeling sets: What we can do and what we cannot do. *ACS Omega, 2*, 2805–2812.

Zheng, G., Han, G., Soomro, N. Q., Ma, L., Zhang, F., Zhao, Y., Zhao, X., & Zhou, C. (2019). A novel computer-aided diagnosis scheme on small annotated set: G2C-CAD. *BioMed Research International, 2019*, 6425963.

Zhou, X., Menche, J., Barabási, A. L., & Sharma, A. (2014). Human symptoms-disease network. *Nature Communications, 5*, 4212.

Zhu, X. W., Xin, Y. J., & Chen, Q. H. (2016). Chemical and in vitro biological information to predict mouse liver toxicity using recursive random forests. *SAR and QSAR in Environmental Research, 27*, 559–572.

Zhuang, J., Tang, X., Du, Z., Yang, M., & Zhou, Y. (2016). Prediction of biomarkers of therapeutic effects of patients with lung adenocarcinoma treated with gefitinib based on progression-free-survival by metabolomic fingerprinting. *Talanta, 160*, 636–644.