

AI in Drug Discovery and Development
Prof. Rajnish Kumar
Dept. of Pharmaceutical Engineering and Technology
IIT-(BHU), Varanasi
Week-10
Lecture-46

Welcome to the course "AI in Drug Discovery and Development." In today's session, we will talk about AI in genomics for personalized treatments. So, by the end of this course, you will be able to understand personalized medicine and explain how genomic insights shape individualized healthcare. Describe the role of genetic mutations in influencing drug metabolism, efficacy, and safety. Understand how genomics contributes to disease stratification, early diagnosis, and targeted therapy. Identify key AI techniques used in genomic data analysis and drug response prediction, as well as explore real-world applications of AI in pharmacogenomics and clinical decision-support systems.

So, let us talk about what personalized medicine actually is. So, personalized medicine, also known as precision medicine, is a medical approach that tailors treatment and prevention strategies to an individual's unique genetic makeup, lifestyle, and environment. So, some of the key characteristics of personalized medicine are that it utilizes genetic information, like genome sequencing, to understand individual variability in disease risk and treatment response. As well as focusing on identifying the right drug for the right patient at the right time, it also shifts focus from reactive treatment to proactive prevention based on individuals' risk profiles.

So, looking at what drives personalized medicine. So, you can see here with this example that it is related to traditional medicine. You will observe that different persons or different people have different therapeutic efficacy. Same dose of the same drug might be like giving a very good response in one person, in another person it might be giving a moderate response therapeutic response and in another person it might be ineffective actually. So, it is not showing any efficacy or response at all.

So, this is due to the difference in the genetic makeup of all these people. So, what personalized medicine offers is that it tries to personalize medicine according to the genetic makeup of the person. so that you get the optimal therapeutic efficacy for the drug, which is individualized and personalized to the patient. So, the one-size-fits-all treatments often fail due to biological differences, and personalized medicine is here to help with that. So, looking at the need for personalized medicine.

So, the first thing is that traditional medicine has a lot of limitations, as we said, and the

one-size-fits-all approach often fails due to genetic and environmental variability. As well as a high rate of adverse drug reactions being observed in the case of traditional medicine, treatment resistance is also being observed in traditional medicine. Many patients experience suboptimal outcomes despite standard treatments as well. Another important issue is the rising complexity of diseases, such as chronic diseases like cancer, diabetes, and autoimmune disorders. These are multifactorial and heterogeneous in nature, and it requires deeper insights into individual biological pathways and molecular drivers to understand the disease pathology.

Another important aspect is genomic evolution. Affordable genome sequencing enables large-scale analysis of individuals, genetic variations, such as the SNP zone mutations, and facilitates the identification of biomarkers and targetable molecular traits. So, if we go back like 20 years, it was very costly to do the genome sequencing, but now it is very fast and very cheap as well. Another important aspect is technological advancements. So, the growth in big data, especially the AI, ML technology advancements and omics technologies, supports the integration and analysis of complex biological data.

And it enables predictive modeling and treatment optimized at the individual level. Another important aspect is the shifting healthcare paradigm. So, we are now moving from the reactive to the proactive approach and also from the population-based to the person-centric care, emphasizing prevention, early diagnosis, and tailored interventions. Okay, let's discuss what a gene mutation is. So, because we are trying to understand personalized medicines or genome sequencing, we need to understand why the same drug gives different effects in different people.

That can happen due to mutations in their genes as well. So, let us see what a gene mutation is. A genetic mutation is a change in the DNA sequence of an organism's genome. So, these mutations can occur naturally or due to environmental factors, such as radiation or chemicals. Since the genome contains all the genetic information for an organism, mutations can alter how genes function, potentially leading to disease or variation in drug response.

So, the mutations can be of two types: one is the point mutation, and the other is the frame shift mutation. So, as I said, what kind of effect these gene mutations can have on drug response? Genetic mutations significantly influence how individuals respond to drugs. So, the key impact includes altered drug metabolism. Mutation in drug-metabolizing enzymes like CYP450 can speed up or slow down drug breakdown. For example, the CYP2C19 mutation reduces clopidogrel activation, lowering its effectiveness.

So, another effect they can have is on the modified drug targets. Mutations in the drug

targets can reduce drug binding and efficacy. For example, EGFR mutations in lung cancer affect the response to TKIs such as erlotinib. So, these tyrosine kinase inhibitors like erlotinib can lose their potency if there are EGFR mutations in the EGFR in the case of lung cancer. And then the mutations can also impair drug transport as well.

So, variation in transporter genes like ABCB1 or SLCO1B1 can alter drug absorption and distribution. For example, SLCO1B1 variants reduce statin uptake, increasing the risk of muscle toxicity. Okay, now we have seen how genetic mutations can alter the drug effect. So, let us see what are the application of genomics actually. So, genomics can help us understand genome biology, host transcriptional responses, and gene expression signatures as well.

So, we can understand the evolutionary history, phylogenetic divergence, transmission dynamics, and outbreaks as well. So, the genome-wide association studies, commonly known as GWAS. So, that can be performed with the help of genomics, and then the candidate gene studies can also be performed, allowing us to understand the genetic diversity and the genomic variations. We can understand the pathogenic molecular physiology as well. We can detect the drug resistance mutations.

We can also, you know, perform the drug resistance surveillance. It can help us with the treatment guidelines, and it can also assist us with drug discovery as well as vaccine development. So, we can perform the whole genome sequencing analysis, and we can understand the host-pathogen genomics and transcriptomics as well. So, ultimately, you know that genomics can help us with a lot of things, and especially, it can help us tailor the treatment to clinically effectively manage the patients. So, let's talk about the role of genomics in personalized treatment.

So, what we can do with it is identify the genetic variants, such as the genomics; it can reveal mutations, single nucleotide polymorphisms (SNPs), and gene expression patterns that influence disease risk and drug response. Then we can develop targeted therapies that can help us match treatments to specific molecular alterations in an individual's genome. For example, how to inform positive breast cancer patients that they can benefit from trastuzumab and that it has been identified that this drug will work in these positive patients through genomic studies. And then it can help us in predicting the drug response. So, genomic data guide dose selection and drug choice, minimizing side effects and maximizing efficacy.

For example, TPMT gene variants affect thiopurine metabolism in leukemia treatment. Then, we can also do early disease detection with the help of genomics. So, genomic screening enables risk prediction and early intervention for genetic disorders and cancers.

And we can do the genomic studies that can help us guide the clinical trials as well. So, we can certify patients based on genetic markers to design genotype-guided trials and improve trial outcomes.

And this we have seen in, you know, the applications of AI in conducting clinical trials as well. Okay, so coming to the data challenges in genomics. Now that we have studied, we have seen how genomics can help us in personalized medicine. So now let us talk about the challenges with the data, actually. So when we talk about these data sets, they actually come from, for example, multi-omics platforms.

So, like genomics, transcriptomics, and metabolomics, they have all these properties. So, like variety, the data is coming not only from one single assay; it can be from heterogeneous sources. So, the data will be of variety, velocity, visualization, volume, and then our veracity, variability, and value. So, you can see here, for example, a single human genome can take up to, you know, 200 gigabytes of raw data. So, now let us see how AI is becoming a game-changer in genomics.

So, it can help us accelerate the data analysis. So, AI processes massive genomic data sets quickly and efficiently, which speeds up variant detection, annotation, and gene expression profiling. It can enhance precision. So those ML models predict disease risk, drug response, and treatment outcomes. Moreover, it identifies clinically relevant patterns that are invisible to traditional methods.

It helps us to enable the predictive modeling as well. AI can forecast disease progression and therapy responses based on genomic signatures and support real-time decision-making in personalized medicine. It can help us with the automatic discovery of new gene disease associations and drug targets that facilitate biomarker discovery for multiomics data. And finally, it can help us integrate the multiomics and clinical data. So, by combining genomics with transcriptomics, proteomics, and patient records, it further helps build comprehensive models for individualized care.

Let us see the big market players in this field. So, few organizations that are working in the direction of integrating AI in genomics are Deep Genomics, Data4Cure, Freenome Holding, Illumina, Sophia Genetics, and Benevolent. And these are not only companies that are working with, you know, AI in genomics. So, there are plenty of them, but these are some of the major players, and you can see here the growth in the genomics market by component, actually from 2021 to 2032. So you can see here we are right now, for example, in 2025, and then you can see that it is exponentially growing not only in the services but also in the hardware and software as well.

Okay, some of the techniques that are being used in genomics. So, we use, for example, machine learning methods like SVM or random forest in variant classification gene-disease association studies. The deep learning techniques, like CNN and RNN, are being used for DNA and RNA sequence pattern recognition. And NLP, those natural language processing techniques are being used for extracting gene-drug relationships from the biomedical texts. So, it can be electronic health records, or it can be, you know, the published literature or the clinical trial data, etcetera.

So, let's talk about the recent investments in AI for genomics. So, Deep Genomics, which is a Canadian company, raised 180 million US dollars in a Series C funding round to advance AI-driven drug discovery, led by SoftBank Vision Fund. And then there are companies like DNA Nexus, Seq1 Genomics, Nonexomics, Variant Bio, and Canadian Genomic Strategies. So, all these companies have received a lot of funding from venture capitalists, or, you know, the investors, to advance the technologies in order to use AI to help with, you know, personalized care and personalized medicine. So, these investments they reflect a global push to harness AI in genomics, accelerating innovation in drug discovery and as well as personalized medicine.

So, let's look at the integration of AI in personalized medicine. So, you can see here on the x-axis, you have the, you know, the advantages like improved drug safety, better patient outcomes, reduced development time, increased efficiency, and enhanced predictive accuracy. And then on the y-axis, you can see the percentage of practitioners reporting benefits. So, you can see that the better patient outcome is one of the major parameters, which is, you know, one of the major advantages that practitioners are observing by integrating AI in personalized medicine. Talking about the applications of AI in personalized treatment, it can help us with disease subtype prediction.

For example, AI classifies complex diseases like cancer into molecular subtypes, and it enables more precise diagnosis and targeted therapies. It can also help us with drug response prediction, like the models that can forecast patient-specific responses to drugs. So, it identifies likely known responders and optimizes the dosing for the treatment. And AI can also help us with, you know, guiding gene editing. So AI enhances CRISPR precision by identifying optimal gene targets and predicting off-target effects, as well as supporting safer and more effective gene therapies.

So these are some of the personalized medicine-focused AI tools, like IBM Watson for Oncology, which uses natural language processing. Some of the applications can recommend cancer treatment using genetic data and clinical literature. And it can integrate genomics, trials, and electronic health records for tailored oncology care as well. Then we have the Tempus, which is, you know, using genomic sequencing plus machine learning.

So, it analyzes clinical and molecular data to recommend cancer therapies, and it is also used for obtaining real-time personalized insights.

Then we have the GNS Healthcare REFS. So, it uses reverse engineering and forward simulation, which can help us with the modeling of biological systems to predict drug response and optimize the drug combos and doses. Then you have the Curate AI, which is using polynomial modeling based on small data and patient-specific phenotypic responses, which is helping us with optimizing drug dosing for individual patients by mapping the relationship between drug dose and patient response, creating a personalized profile that adjusts dynamically. And then you have the TransDTI. This is a transformer-based language model that estimates drug-target interactions, which are crucial for personalized treatment approaches.

PHARM BERT, which is a domain-specific BERT bidirectional encoder representation from the transformer model, enhances drug safety monitoring and interaction analysis. Talking about one of the latest developments, which is alpha missense. So, it is DeepMind, which is a Google subsidiary. So, DeepMind AI for genetic mutation prediction. So, AlphaMissense can predict whether a missense mutation, which is a single-letter DNA change, is likely to cause a disease or not.

The characteristics of this tool are that it covers 71 million possible missense variants. And then you know it covers like 89 percent prediction coverage with 90 percent accuracy, and it classifies variants as, for example, 57 percent probably benign and 32 percent probably pathogenic. And how it is impacting genomics is that it can accelerate rare disease diagnosis as well as expand variant classification from 0.1 percent, which is traditional, to 89 percent. And then it speeds up research by spotlighting the high-risk mutations, which are the mutations that will definitely lead to a disease.

And then we will have a look at the curated AI workflow. So, how it works is that it collects the patient data, and then this data is used to generate the personalized profile. And then analysis is being done where the patient response predictions are made. So, based on this response prediction, a dose is being recommended that is okay for this patient if they have the optimal genomic profile.

So, this patient will respond to the normal dose. Based on the recommended dose, the guided dosing is then done. And if the guided dosing is, you know, giving the desired therapeutic response or not based on that, this model can be recalibrated using the latest patient data if necessary. So, this is how the curate AI workflow is working to, you know, guide the dosing. Okay, let's come to the applications of AI in pharmacogenomics. So, pharmacogenomics is the study of how an individual's genetic makeup influences their

response

to

drugs.

So, it combines pharmacology and genomics to develop personalized medicine that is safer, more effective, and tailored to each person's genetic profile. So, the AI technologies enhance this field by analyzing complex datasets to guide safer and more effective treatments. Like it can help us with drug response prediction, those ML models predict how individuals metabolize or respond to drugs, such as warfarin or clopidogrel. And then it can help us with the variant impact assessment, like AI can identify genetic variants affecting drug metabolism, enzymes like CYP2D6, CYP2C19, and SLCO1B1.

And then it can help us with preventing adverse reactions. Those NLP and ML models can scan the biomedical literature to flag variants linked to drug toxicity. And then I can definitely help us with the dose optimization, as the AI can recommend personalized dosing strategies based on genomic and clinical profiles. So, let us see a case study in the field of pharmacogenomics. So, it guided the Warfarin dosing in personalized drug discovery. Warfarin is, you know, a commonly prescribed anticoagulant used to prevent blood clots.

However, the problem is that having a very narrow therapeutic index means too much can cause bleeding and too little of the drug can cause clots. So, the genetic variability among individuals makes standard dosing risky. So, there is up to a 20-fold variation in dose requirements among patients. So, you can see that there is a lot of variation among the patients. Therefore, we need to identify how much dose will be needed for a specific person.

So, that is the kind of personalized, you know, medicine. So, around 50 to 20 percent of the patients experience bleeding or clotting due to improper dosing. Let us see what those genetic influences are on warfarin response. So, there are two key genes that impact the warfarin dosing. One is the CYP2C9, which affects how fast warfarin is metabolized, and there are variants, actually CYP2C9-2 and CYP2C9-3, which show slow metabolism; therefore, the carriers require a 30 to 90 percent lower dose to avoid overdose.

So, there are people who have these variants of CYP2C9. So, they, you know, metabolize warfarin at a slower rate than the normal, you know, CYP2C9 enzyme. So, in that case, they need at least a 30 to 90 percent lower dose to avoid an overdose. And then there is a VKORC1, which affects warfarin sensitivity. So, the 1639G variant reduces enzyme activity. So, individuals with AA genotypes, common in Asians, need a 45 percent lower dose than the GG carrier.

So, then coming to the dosing algorithm and clinical validation. So, the International Warfarin Pharmacogenomics Consortium created a dosing algorithm combining the

genetic data, which included CYP2C9 and VKORC1, and the clinical factors like age, weight, and other medications. However, the limitations of this dosing algorithm were that it was effective in Caucasians, where it explained 50 percent of the dose variability. However, it was less accurate in Asians and Africans due to the missing population-specific variants, as those variants are more prevalent in African or Asian populations. So, that information was not available. So, then a trial was done, the gift trial, and the genetic informatics trial.

So, to test whether genotype-guided dosing improves outcomes, in this trial, 1,650 post-surgery patients were included, and the results indicated that adverse events were reduced. So, 10.8 percent in the genotype-guided group versus 14.7 percent in the standard group represents a 27 percent relative reduction.

And then there was better control. So, the time in therapeutic range (TTR) was 67.4 percent in the genotype-guided group versus 62.7 percent in the standard group. So, AI-driven enhancements in the warfarin doses were performed. So, the machine learning was integrated using NGS for rapid genotyping in less than 24 hours, and the AI models trained on diverse populations improved dose predictions by 12 to 18 percent in ethnic minorities as well.

So, the clinical implementation, for example, at London Health Science Centre cut Warfarin-related hospitalizations by 22 percent using AI-guided pharmacogenomic protocols. So, you can see how effective these algorithms can be in deciding the clinical dose and optimizing the dose of a drug for the patients. So, talking about the challenges and solutions, the ethnic disparities were a major challenge where early models underperformed in non-European groups. So, the solution was to train AI models on diverse datasets using federated learning, and federated learning was used because sometimes databases are not ready to share their data because it contains a lot of sensitive information. So, in the case of federated learning, the data is not being shared, but the features of the data are being shared, and then that can be used for developing a model.

And then it included CYP4F2 as well, which explains 7 to 9 percent of variability in East Asians. So, that also help in you know in improving the dose prediction. So, talking about the cost versus benefit, the genotyping cost roughly 200 dollars per patient, but it led to 4000 to 6000 in annual savings due to the reduced hospitalization. So, first they need to do the genotyping, and then they can decide the dose, but that genotyping cost was saving a lot of money in the form of hospitalization.

Okay, so let's talk about the AI in clinical genomic decision support. So AI can enhance clinical decision-making by analyzing genomic data to guide diagnosis, prognosis, and

personalized treatment planning. So it can support evidence-based as well as personalized care. So, the key role AI can play here in this case is variant interpretation. AI helps classify genetic variants as pathogenic versus benign from the sequence data. And then it can help with the risk prediction, where ML models assess genetic risk for diseases, such as cancer or cardiovascular conditions.

And it can help us with therapy matching, where it integrates genomic profiles with clinical data to recommend the most effective treatments. and the knowledge integration where NLP tools extract relevant insights from biomedical literature and databases to support clinical choices. So, if we look at this, you know, AI decision support system. So, you can see that it can help us with all sorts of, you know, decisions like the preventive, predictive, participatory, and personalized. Where is the predictive collecting and analyzing of clinical and non-clinical data as risk factors for disease development? And then in the preventive context, where you know everyday life monitoring and self-management technologies for proactive preventive measures are being used, this leads to the diagnosis support system.

And then we have the virtual care system where you have personalized, tailored medical plans for individual patients based on comprehensive medical data. As well as enhancing patient engagement and adherence through ongoing monitoring and support in the care process. Okay, then another example is the use of AI in rheumatology, where it is again an example of precision medicine. So, AI is transforming rheumatology by enabling accurate diagnosis, patient stratification, and tailored treatment for rheumatic diseases like rheumatoid arthritis. So, some of the key innovations are that AI platforms for rheumatoid arthritis quantify radiographic damage and optimize treatment selection.

And then Stanford's revamp integrates genomics and proteomics to match autoimmune patients with the right therapies. And then you have the ML models developed by Guan et al. So, they predict RA progression and respond to anti-TNF alpha therapy using genomic and clinical data. So, the real-world impact is that this RSF model predicts RA patients' mortality risk using clinical or serological data and then the supervised ML models. So, they forecast methotrexate response based on the clinical and pharmacogenomic inputs.

Okay, let us see what the future perspectives are of using AI in genomics. So, it can enhance, you know, multiomics integration. So, AI will increasingly integrate genomics, transcriptomics, proteomics, and metabolomics to create comprehensive patient profiles for highly precise treatment strategies. And then it can help us with real-time clinical decision support. So, that is, you know, the future; actually, the AI-driven tools will offer instantaneous genetic insights at the point of care, enabling dynamic individualized treatment plans during routine clinical workflows. And federated learning and data privacy,

where privacy is, you know, a matter of real concern.

So, privacy-preserving AI approaches like federated learning will allow secure collaborations across institutions and countries without data sharing. And cost-effective genomic medicine with automation and AI-based diagnosis will reduce the cost of sequencing and interpretation, making personalized treatment accessible to broader populations. So, coming to the summary, the genomic insights form the cornerstone of precision medicine, enabling tailored intervention based on individual genetic variability. Integration of high-throughput gene sequencing and bioinformatics allows for the identification of disease-associated variants, guiding targeted therapies and early diagnosis. Personalized treatment strategies benefit from pharmacogenomic profiling, minimize adverse drug reactions, and enhance therapeutic efficacy.

And AI plays a transformative role in genomic analysis, streamlining data interpretation, variant classification, and predictive modeling. And clinical decision support systems powered by AI align genomic data with patient records to recommend personalized treatment pathways. And ongoing advancements in AI and multi-omics integration continue to drive the evolution of individualized care, setting new benchmarks in precision healthcare. So, with that summary, I have an open question for you. If AI can integrate genomics, lifestyle, and environmental data to design truly personalized treatment, could we envision a future where AI-curated prescription genomes are custom-written for each individual, reprogramming gene expression in real time to prevent or cure disease before symptoms ever emerge? So, just think about it.

And I have some suggestions, you know, as some papers that you can go through to increase your knowledge in this area. And with that, thank you.