AI in Drug Discovery and Development

Prof. Rajnish Kumar
Dept. of Pharmaceutical Engineering and Technology
IIT-(BHU), Varanasi

Week-04
Lecture-17

Welcome to the course "AI in Drug Discovery and Development." So, now that we know about the drug targets. So, let us take a look at drug target identification and validation. So, we will take a look at the basics of drug target identification and validation in this session. So, by the end of this lecture, you will be able to understand the importance of target identification in drug discovery. Learn different methods for target identification, such as genetic studies, functional genomics, and omics approaches.

Also appreciate the role of target validation in ensuring therapeutic efficacy and safety, as well as explore the future trends in target identification and validation. So now that we know there are multiple kinds of drug targets, such as enzymes, GPCRs, ion channels, protein-protein interfaces, nucleic acids, and transcription factors. So let us have a look at what makes a good drug target. So, the first thing that every drug target should have is an assessment of whether the target has a proven disease-modifying function or not.

as well as modulation of the target is less important under physiological conditions. For example, if the drug target is involved in many physiological functions and if we modulate that with a small molecule or a drug. So then, that will lead to a severe side effect because we are altering the normal physiological functions as well. So, we need to validate the drug target based on the disease as well. So, the second property of a good drug target is druggability.

Means that if the druggability is not precedented, a three-dimensional structure for the target protein or a close homolog should be available for a druggability assessment. And then the target should have favorable assayability enabling high-throughput screening, which means that we need to know the 3D structure of the target. so that we can see how the drugs we are developing for this target are interacting with it and then altering its function. So, the target should be favorable to the assay development; only then will we be able to perform the high-throughput screening to identify hit compounds. And then the third important aspect of a good target is safety and efficacy.

So, the target expression is not uniformly distributed throughout the body. Again, if the target is developing a drug for that target, it will lead to severe side effects because it will affect it everywhere in the body. And then a target or disease-specific biomarker should

exist to monitor the therapeutic efficacy of the target, as well as a favorable prediction of potential side effects according to the phenotypic data. For example, knockout mice or a genetic mutation database so that we can see whether altering this target is leading to the actual desired therapeutic effect or not. If you look at this small study where they performed an analysis of clinical trial failures from 2013 to 2015.

So we can see that efficacy remains the primary issue for failure in both phases, with no improvement in the proportion of phase 3 failures due to efficacy compared to phase 2. So, in phase 2, around 48 percent of the failures were due to insufficient efficacy. And 25 percent of the Phase 2 failures were due to safety concerns. Combining, 73 percent of all phase 2 failures were due to efficacy or safety issues. And if you trace it back to the drug targets, we can see that both efficacy and safety are actually related to the target.

So, if the target is, as I said, spread throughout the body, that will lead to safety issues because it will affect the normal physiological functions and will have severe side effects. Again, in phase 3 failures, 55% of phase 3 failures were due to insufficient efficacy, and 14% of phase 3 failures were due to safety concerns. So, overall, if you look at this analysis, the major reason for the failure of the drugs in phase 2 and phase 3 clinical trials was efficacy and safety. So, what matters in identifying a good target is that we need to have access to models and technologies by which we can evaluate, identify, and validate that target. So, we need to have an understanding of the basic molecular mechanism of inhibition of that target, and we need to have a basic understanding of the disease as well.

So, if we combine all these and fulfill all three criteria. So, what we will have then is a good drug target. Coming to target identification. So, this is the first step in discovering a new drug, and what we do here is identify the function of a possible therapeutic target, and we also identify the molecular mechanism in the disease. So, it involves the recognition and validation of biological targets that interact with bioactive compounds, facilitating the understanding of disease mechanisms and the development of therapeutic interventions.

So, this consists of both the molecular target as well as either the protein, receptor, or ion channel that we have discussed earlier, which a drug is designed to affect. As well as it is related to the asset, the asset is a drug or therapeutic candidate that is under development and testing for clinical use. We need to identify both ways through the perspective of the asset, as well as how this molecule is exerting its therapeutic activity. And we need to see it through the molecular target as well, like how that particular target is being affected by those inhibitors. So, if you look at the evolution of target identification, back in 1991, protein identification was facilitated by photoaffinity labeling, and this crucial research led to the identification of several drug targets.

And then, in 1994, the mTOR discovery was made, which is a target of rapamycin. It further advanced our understanding of cellular growth and metabolism. And then in 2002, the introduction of SILAC, which is stable isotope labeling by amino acids in cell culture for precise protein quantification in proteomics, was made. And then in 2003, the completion of the human genome project was achieved, and simultaneously, the pharmacophore screening for target identification was also reported. And then in 2004, the development of high-throughput target docking for computational drug discovery.

As well as the identification of EGFR mutations in known small cell lung cancer, leading to the discovery of targeted therapies. And in 2005, the application of genome-wide association studies (GWAS) was utilized to uncover genetic variants associated with the diseases. And in 2006, the introduction of structural similarity assessment for predicting drug-drug interactions based on molecular structures was reported. And then, in 2008, affinity-guided catalyst chemistry was applied in drug discovery, focusing on kinase targets. So, in 2009, ligand-directed chemistry was introduced for ligand discovery and optimization in drug development.

followed by the development of CRISPR-Cas9 technology in 2012, which revolutionized functional genomics and target validation. So, in 2014, the development of the original GAN, Generative Adversarial Network, and DL-based QSAR, Structured Activity Relationship, methods for drug discovery were revealed. And in 2017, the disease-driven target identification using deep learning was also reported. So, the year 2020 witnessed a very important development in which DL-based target identification was applied to COVID-19 research, accelerating therapeutic development against this deadly disease. And in 2022, the entry of the first AI-designed drug into clinical trials happened.

which was identified by in silico medicine. Furthermore, the identification of AI-derived novel targets for ALS treatment was also reported again by in silico medicine, and these will be discussed in the coming sessions. And then, in 2023, the integration of advanced AI chat functionality with knowledge graphs for enhanced target identification was also reported. So, you can see that target identification has evolved from the classical methods to the AI and deep learning-based methods. So, if we talk about the target identification methods for molecule targets, there are genetic studies such as genome-wide association studies known as GWAS, and then genomic sequencing like whole genome sequencing and whole exome sequencing.

And then we have functional genomics, like CRISPR-Cas9 gene editing or RNA interference, RNAs. And then we have omics approaches as well, where we utilize transcriptomics, proteomics, metabolomics, and other omics as well. And then for the asset, we usually have this phenotypic drug discovery. And once we identify a molecule that is

working in the phenotypic screening, we do not know how it acts. So, then we go back and try to deconvolute the target.

So, by using techniques called target deconvolution, we try to identify the targets of those molecules that are showing effects in the phenotypic drug discovery screening. And then we have target-based discovery, where we know the structure of the target, and then we try to identify inhibitors for that target. So, if you look at the GWAS, the GWAS analyzes genetic variations such as single nucleotide polymorphisms across the entire genome to identify associations between specific genetic variants and disease susceptibility. So, what we do in this study is collect DNA samples from cases and controls, followed by performing high-throughput genotyping. Apply quality control measures and use statistical analysis to identify genetic variants significantly associated with the trait or disease.

So, the genetic variants identified through GWAS can point to potential target genes or pathways involved in disease development. So, one of the successful examples of using GWAS studies is the identification of the APOE gene as a well-established genetic risk factor for Alzheimer's disease, which was identified through GWAS studies. So, you can see that we can identify all those kinds of, you know, disease-associated single nucleotide polymorphisms followed by the genes that are altered in those diseases and the causes of those diseases. And then another technique is called genome sequencing, where we use high-throughput DNA sequencing techniques such as whole genome sequencing (WGS). Our whole exome sequencing enables us to identify the genetic mutations, including rare variants, in individuals or patient populations.

So, the WGS identified mutations in the CFTR gene associated with cystic fibrosis as one of the examples of the success of using genome sequencing in identifying the drug target for this disease. And then we also use functional genomics. So, functional genomic studies involve manipulating genes in model organisms or cell lines to assess their impact on disease-related phenotypes. So, the functional genomic experiments such as knockout mice or RNA interference screens validate the functional importance of specific genes and their suitability as therapeutic targets. This CRISPR-Cas9 technology allows researchers to precisely edit specific genes or genomic regions, providing insights into gene function and potential therapeutic targets.

So then we have omics approaches like transcriptomics, which identify genes with altered expression. Either the genes are upregulated or downregulated in disease using techniques like RNA sequencing and RNA-seq, while also pinpointing potential therapeutic targets such as identifying oncogenes in cancer. And then we use proteomics to analyze the protein abundance, modifications, and interactions using tools like mass spectrometry. which help us to uncover the dysregulated protein involved in disease pathways, such as the discovery

of amyloid beta in Alzheimer's disease or the tau protein in Alzheimer's disease. And then there is a term called multi-omics, which refers to this multi-omics approach where we combine multiple omics technologies like genomics, transcriptomics, proteomics, and epigenomics.

And other omics data to identify molecular targets by providing a comprehensive view of disease mechanisms and uncovering complex interactions with very high precision. So, those were techniques for identifying the molecular target. So, if we talk about the target identification for asset, so how an asset means a molecule which has shown activity in phenotypic screening that is showing its effect. So, how do we determine that? So, there are basically two aspects: one is phenotypic drug discovery, and the other is target-based drug discovery. So, the advantage of phenotypic drug discovery is that we do not need to know the molecular target of a disease, nor do we need to know where this molecule is going to bind or which enzyme it is going to block.

So, we directly throw our compound into the assay screen and then we determine whether it is showing any effect in the screens or not. For example, if you wanted to see whether a compound is showing anti-cancer activity, you can just treat the cancer cell lines. Or the cancer tissue with those compounds and see which of those compounds shows activity, inhibiting the growth of cancer cell lines. So, in that case, we will not be able to tell which molecular target those molecules are binding to, but we will be able to determine that later by using the target deconvolution technologies. That there is possibility to discover molecules with unique mechanism of action and then we can develop the novel biological discovery as well.

So, the disadvantages are low throughput because we cannot we cannot assay a million or millions of compounds using the phenotypic screening because it is time intensive and it is like very costly. And then another disadvantage is low stability and robustness of the assay system. And sometimes the target deconvolution and the mechanism of action can be challenging. So, because it is so complex that we will not be able to know how exactly this molecule is showing the activity. So, the advantages of target-based approaches are that they are high throughput.

So, you can maybe screen millions of compounds without any hustle. And then there are robust screening cascade has been established and also structure based approach is also possible. So, you can use the three-dimensional structure of the receptor or the enzyme of course of the drug target and then you can utilize all those structure-based method for determining how for identifying the lead compounds or hit compounds. And then the disadvantages are that the hit compounds may not show efficacy in cell-based assays. Sometimes in purified enzyme-based assay or in purified system, so those inhibitors, they

are showing very good activity.

But as soon as you treat the cells or you go to the in vivo studies, they do not show any activity. And there is a reason behind that because when you are treating a cell or an animal with the molecule So, that molecule may not be soluble, that molecule may not be permeable to the cell membrane, that molecule may not be bioavailable. So, there are lot of you know hurdles in between. So, they work very nicely in the purified system, isolated system, but they might not have effect in the cell-based assays or in the animal models. So, then there is the shortage of druggable target molecules as well and then if we talk about the material which is being used for phenotypic drug discovery are the cells, tissues or organ and for the target-based discovery we use the common protein or the cells.

So, then we come to the target validation. So, once we have identified a target, now it is very important to validate this target. So, this is, this confirms the biological relevance and therapeutic potential of the identified targets and then where we ensure targets are druggable and linked to disease mechanism and here druggable means that we are able to develop small molecule or the biologicals who can engage those targets and lead to the desired therapeutic effect. And it is very critical to reduce clinical trial failure and improve drug efficacy in long run. So, there are in vitro validation methods like we use the cell lines, primary cells or tissue cultures to assess target function and there are techniques which include gene knockdown, RNAIs, gene knockout, CRISPR Cas9 and pharmacological inhibition assays. For example, the CRISPR knockout of KRAS gene in cancer cell lines validated its role as a key oncogene.

So, this was only possible due to this CRISPR technology that the knocking out of the KRAS gene validated its role in the, its key role as an oncogene. And then there are in vivo target validation method which uses genetically modified animals like knockout mice, transgenic mice or genograph models to confirm target function in disease context. So, what we do is we evaluate the therapeutic efficacy, safety, biological relevance in whole organisms instead of in the isolated system. So, mouse xenograft models which validated BRAF V600E mutation, inhibition efficacy by Vemurafenib in melanoma. So, you can see the success of using these mouse xenograft models.

And these are mostly being used nowadays for doing the drug discovery. So, nowadays there are like humanized mouse models or xenograft models which are being used to validate the drug targets for development of the drugs against different diseases. And then there is a biomarker-based target validation as well which utilizes measurable biological indicators, those are called as biomarkers to confirm target engagement and therapeutic response. So, these biomarkers can be genetic, this can be protein based or this can be imaging based indicators which reflects disease progression or drug response. So, for

example, the reduction in LDL cholesterol level validated LDL receptors as targets for statins in cardiovascular disease.

So here the LDL cholesterol level, low density lipoprotein, so this LDL, so they act as a biomarker for measuring the cardiovascular disease. And then there are some integrated approaches like integration of multiple validation approaches provide robust evidence for therapeutic targets. And recently a lot of advancement has happened in target validation as well. So, like the advancement such as organoids, patient derived xenografts known as PDX, single cell analysis and AI driven biomarker discovery, it enhances validation accuracy. And comprehensive validation of PD-L1 as an immunotherapy target involved in vitro assays, animal models and biomarker based clinical studies leading to successful immunotherapies like pembrolizumab.

So if you look at the summary so the target identification involves pinpointing disease related molecular targets using genomics proteomics and functional studies and the target validation confirms the therapeutic relevance and drug ability of these targets through in vitro in vivo and biomarker based approaches. Both these processes are crucial for ensuring effective and safe therapies while reducing drug development failures. And the advances like multiomics, AI, and advanced experimental models, these are enhancing precision in drug discovery. In the end, I have an open question for you. How can integrating multiomics data like genomics, transcriptomics, proteomics improves the precision of identifying and validating therapeutic targets for complex diseases like cancer or neurodegenerative disorders? And you can go through these papers.

So, these are really useful papers. You can go through these and then you can learn more about target identification and validation. And with that, thank you so much.