Big Data & Drug Discovery

By Peeyush Prasad, MSc

Contributing Editor

High-Throughput Strategy for Drug Discovery

How Integration of Data Leads to Drug Discovery

To understand any scientific phenomenon one has to apply careful strategy of designing and proving hypothesis. In many areas of science including biology we want not only to understand any phenomenon but our main aim is to use that understanding for human use. Exploring the disease at cellular level requires the information about all the cellular components. Drug discovery is a complex process from finding lead compound against a particular molecular target to developing safe drug for use. To find the drug target at molecular level one can either use the existing knowledge to test the hypothesis or one can generate the data at global level for finding all the alteration or aberration. Diseases are usually caused by change in the molecular dynamics which can be genetic or epigenetic. Major components of these changes are genes, transcripts, proteins and metabolites.

Approximately, 25000-30000 genes are present in a human cell. These genes express differentially in different types of cells which lead to specific function or phenotype. A slight change in these genes may have detrimental effect on the cell. Many diseases are complex in nature and it is difficult to pin point a specific cause. To overcome these challenges, high-throughput based approaches are employed. Approaches such as genomics, transcriptomics, proteomics and metabolomics are employed to understand disease or to find a drug particular target. These approaches use the high-end technology which requires highly trained man power but they take less time and give definitive answer in a particular context. Integration of artificial intelligence and high-throughput data has made the process more refined and accurate. Further single cell omics analysis and system biology approach has been developed for better understanding the data. Figure 1 depicts the integration of different data for drug discovery.

Different High-Throughput Approaches for Drug Discovery

Genomics seek to understand gene and its regulatory element. Genomics profile gives

the information regarding mutations, single nucleotide polymorphism (SNPs) and information regarding regulatory elements. Diseases such as cancer are complex disease and there are several types of mutations present in a tumor. Mutation in the genes leads to change in several signaling and metabolic pathways. These changes may cause the survival and progression of cancer cells. Similar is the case with other disease such as diabetes. Therefore, finding the dysregulated molecules at gene level may help in finding the factors responsible for disease and its progression.

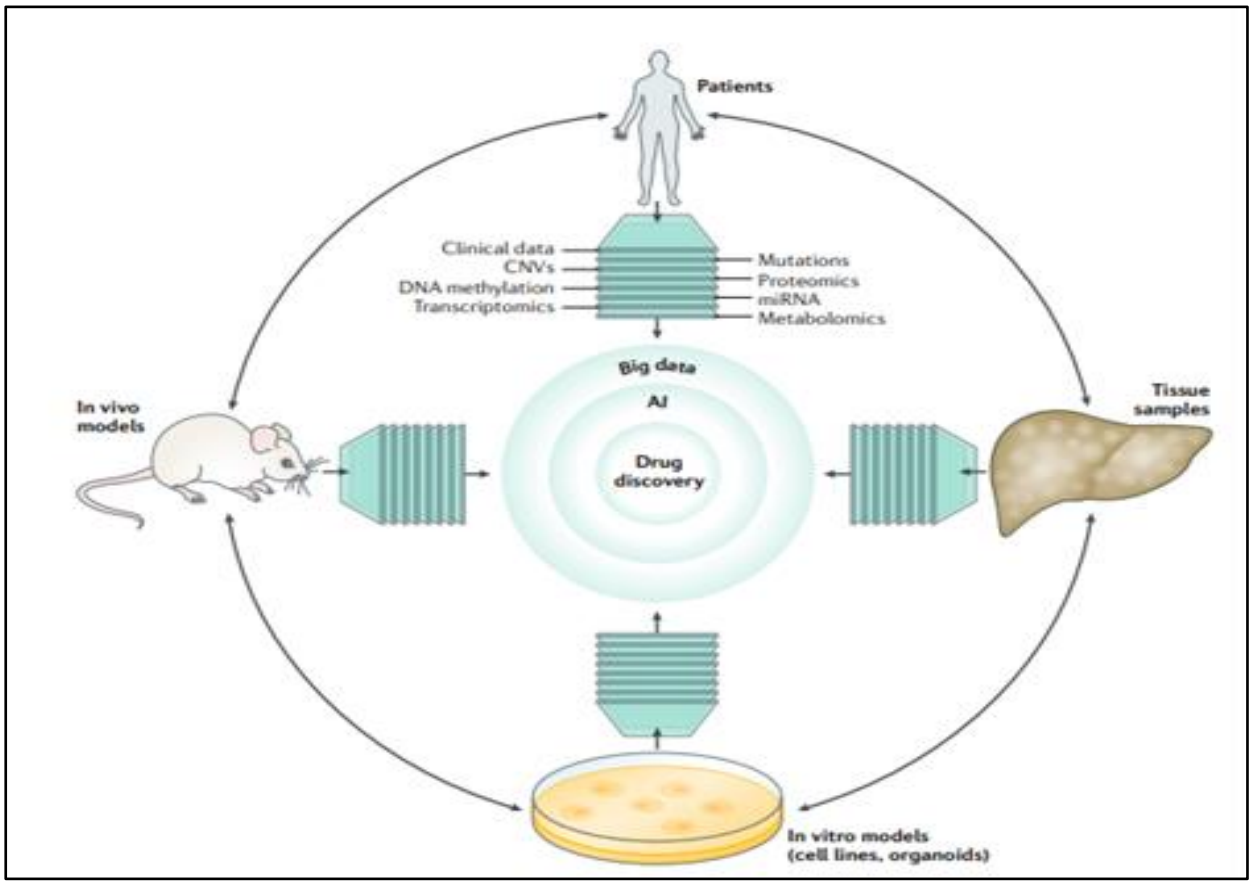

*Figure 1. Translational research and big data. Translational research comprises four main components: patients, tissues, in vitro models (cell lines and organoids) and in vivo models. Each component can be characterized by different molecular modalities (such as genomics, epigenomics and functional genomics). Artificial intelligence (AI) can be used to improve the insights from big data by delineating differences and similarities and further facilitating efficient therapeutic discovery. CNV, copy number variation; miRNA, microRNA [**Source:** Nature Reviews, Chen B et. al., 2020].*

Further mutational analysis also helps in identifying the mechanism of drug resistance. Mutational landscape of cancer cells vary from patient to patient [2]. Therefore, genomic profile of a patient can help in designing the personalized medicine. Transcriptome profiling of a disease gives the idea about the expressed genes. Only limitation of RNA based analysis is not all the mRNAs get translated into the proteins and hence do not contribute to the phenotype. One of the advantages of trascriptome based analysis is it gives the idea about the regulatory RNA such as miRNA and lncRNA.

Company	Function	Link
Agilent Technologies	Working in the field of biopharm cancer research and cell analysis along with diagnostics, food and energy.	https://www.agilent.com/
Danaher	Genomic company working in the field of healthcare, diagnostics and life sciences.	https://www.danaher.com/
23andme	Genetic testing and analysis	https://mediacenter.23andme.com/
Verge Genomics	Uses machine learning and human genomics for drug discovery.	https://www.vergegenomics.com/
Cytomyx	Company works on comparative proteomic analysis of different samples for drug discovery.	https://cytomx.com/
Evotec	Provide advanced technology platform for protein expression analysis and modifications.	https://www.evotec.com/en

Table 1: *Top companies working in the field of high-throughput technology.*

Proteins and metabolites are major factors responsible for cell behavior and phenotype. Proteomics based analysis helps in identifying the drug target. Mutations at gene level are often reflected in protein structure which may cause its activity to increase or diminish. Further, change in protein level is often observed in diseases. Proteome based analysis reveal expression level of proteins and bioinformatics based analysis of the data can help in identifying the drug target. Several diseases are the product of error in metabolism and metabolomic profile can give us idea about the indicators of cellular state

such as bio-energetic pathways, toxicity level, anabolic pathways and pathways associated with drug metabolism. Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) tests are usually performed to predict the effectiveness and safety measures associated with the drug [3]. Metabolic profile of a disease patient can help us in designing drug which is not only effective but safe also. Table 1 mentions some of the top companies working in the field of high-throughput technology.

Concluding Remarks and Future Perspective

High-Throughput study is the future for drug discovery science. More and more companies are now employing these technologies for finding better drug. Several high-throughput studies have identified the overlapping molecular mechanism between two different diseases [4]. This disease-disease relationship can help in repurposing the drugs where one drug can be used for more than one disease. On the other hand multi-omic analysis will help in identifying the exact drug target and its effectiveness. Some of the technological challenges still we need to overcome such as development of platform with high sensitivity and specificity. For example, false identification in high-throughput study is one of the major problems. Easy to use pipeline needs to be developed for analysis of large data sets. Integration of AI, high-throughput data and chip based technology can further enhance and accelerate the drug screening process.

References

1. Chen B, Garmire L, Calvisi DF et. al., Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol. 2020 Jan 3.
2. Appenzeller S, Gesierich A, Thiem A et al., The identification of patient-specific mutations reveals dual pathway activation in most patients with melanoma and activated receptor tyrosine kinases in BRAF/NRAS wild-type melanomas. Cancer. 2019 Feb 15;125 (4):586-600.
3. Guan L, Yang H, Cai Y et. al., ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness. Medchemcomm. 2018 Nov 30;10 (1):148-157.
4. Menche J, Sharma A, Kitsak M et. al., Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015 Feb 20;347 (6224):1257601.