Open Source Drug Discovery: From Genome Sequence to Lead Identification

R. Vashisht[1,2], D. Kaur[1,3], OSDD Consortium[1], S. K. Brahmachari[1,2,3]*
[1]CSIR-Open Source Drug Discovery, New Delhi, India
[2]Academy of Scientific and Innovative Research, New Delhi, India
[3]CSIR- Institute of Genomics and Integrative Biology, New Delhi, India
Mailing address, skb@igib.res.in
* Corresponding Author

Keywords: *Open Source Drug Discovery, Innovation, Tuberculosis, Crowdsourcing, Cheminformatics.*

Abstract

The advent of next generation sequencing technologies, high-throughput measurements of omics data and 3D structures along with clinical phenotype association studies has created a data deluge. However, explosive growth in biomedical data generation has not yet translated to proportionate increase in clinical returns. Its full impact on rapid diagnosis, identification of lead molecules with least in vivo toxicity for novel therapeutics and rapid cure for long standing neglected diseases such as Tuberculosis is yet to be realized. Towards this, an Open Source Drug Discovery (OSDD) platform (www.osdd.net) was created for building a systems level understanding of Mycobacterium tuberculosis (Mtb) genome to identify critical novel drug targets. The unstructured knowledge was collated by developing a crowdsourcing approach for genome annotation, resulting in 99.5% of functional annotation in contrast to the 52% reported earlier. The largest manually curated interactome of Mtb was built to identify putative drug targets. Its metabolic complexity was further mapped by accounting for 961 metabolites, 890 protein-coding genes involved in 1152 reactions in 50 pathways by developing a Systems Biology Spindle Map (SBSM). The computational simulation of SBSM in Mtb revealed various essential proteins that are required for the optimal growth of bacteria under myriad conditions. They were assessed as important drug targets by comparison with the human genome and microbiome at sequence and structure level. The identified putative drug targets were further analyzed for their conservation across the 1849 Mtb clinical isolates. Targets with least or no variation across the entire 1849 genomes were evaluated at their PDB structure level. Leveraging the active site of these proteins, and utilizing structure based ligand design approach; set of experimentally validated anti-tuberculosis compounds were identified as drug candidates based on docking and MD simulations. Thus, we have developed a novel systems biology platform; integrating omics data, computational simulations, 3D structure, and available experimental evidences for drug discovery.