Drug Repurposing Guidebook

Building Block I441

This document defines the content of the FACT SHEET to be created for each identified tool, incentives, initiative or practice (the Building Block) introduced by public bodies or used by developers to expedite drug repurposing in Rare Diseases (RDs).

ITEM	DESCRIPTION
Building Block (BB) Title	Machine learning & data mining
References	Bio.tools https://bio.tools/ doi: 10.1093/nar/gkv1116 Drug Repurposing Knowledge Graph https://github.com/gnn4dr/DRKG/ https://arxiv.org/pdf/2010.11367.pdf
Description	Bio.tools is a general repository for software tools, databases and services for bioinformatics and the life sciences. Relevant machine learning methods can be retrieved through this search: https://bio.tools/t?topicID=%22topic_3474%22. Bio.tools uses terms from the EDAM ontology to describe the features of the tools in the repository, to increase findability. Deep learning methods, and more specifically graph convolutional networks, constitute a novel class of machine learning methods that appears particularly useful for predicting novel drug repurposing candidates. These methods use existing information from expert-curated or non-curated (primarily obtained from text mining the scientific literature) sources about molecular defects observed in rare diseases, disease phenotypes and drugs proven to reverse disease phenotypes captured in a drug repurposing knowledge graph. The Drug Repurposing Knowledge Graph (DRKG) is one of the most extensive knowledge graphs available for drug repurposing studies. Although initially focused on identifying drug repurposing candidates for Covid-19, the graph can be used for other diseases as well. DRKG incorporates DrugBank, Global Network of Biomedical Relationships (GNBR), Hetionet, STRING, IntAct, and DGIdb.

ITEM	DESCRIPTION
Category	Compound and network databases and tools to use them
Type of BB	Development practice
Geographical scope	International
Availability	Open Access (code and repository)
Scope of use	The problem of predicting drug repurposing candidates can be defined as a machine learning problem where new links between drugs and diseases are predicted from a knowledge graph through training of a graph convolutional network algorithm on known drug:disease relationships. This is called link prediction. The algorithm is trained on true positives (drugs known to treat diseases), but also requires a set of true negatives. In practice, these are difficult to find because of the underreporting of drugs that are not effective in treating a disease, and therefore a subset of drug:disease combinations that are not yet in the knowledge graph is taken as true negatives. Workflows and service bundles that encompass different steps in the drug repurposing could be a useful step. This was successfully done for other pressing problems like the pathogenicity assessment of genetic variants.
Stakeholders involved	Preclinical researchers from academia and industry
Enablers/ Requirements	Application of this type of advanced machine learning methods requires bioinformatics and programming skills (primarily python and use of PyTorch libraries). Training in deep learning is also required.
Output	A list of drug repurposing candidates for a specific disorder that can be tested in further wet-lab studies.
Best time to apply and time window	Discovery of drug repurposing candidates

ITEM	DESCRIPTION
Expert tips	A comprehensive evaluation of the predictive performance of this type of deep learning algorithms is still lacking, in particular for rare disease, i.e., it is unknown how many of the predicted drugs can be validated in preclinical and clinical studies. Current drug repurposing knowledge graphs are mostly generic and may not contain information about the rare disease of interest or lack information on the context of the disorder (phenotypes, affected tissues, affected molecular pathways). When predicting drug repurposing candidates for a given disorder, it is wise to add information about the specific disorder to the knowledge graph and to keep only the parts of the knowledge that are relevant in the context of the disorder. Several additional resources can be included in the building of knowledge graphs, including resources specified in building blocks 131 (Drug databases), 105 (chemical compound databases), 106 (network databases)