Accelerating Drug Discovery in Rare and Complex Diseases

Shima Dastgheib, Craig Webb, Qiaonan Duan, Rowan Copley, Gini Deshpande and Asim Siddiqui

NuMedii,Inc. , San Mateo, CA, USA
{shima.dastgheib, asim.siddiqui}@numedii.com

Abstract. We report the adoption of Semantic Web technologies by NuMedii to lay the foundation for accelerating drug discovery in complex and rare diseases.

Keywords: Drug Discovery · Fibrotic diseases · RDF graph database.

1 Overview

Despite the incredible advancements in prognosis, diagnosis, treatment and management of many chronic diseases, there remain thousands of complex or rare conditions that are still challenging to manage and are usually incurable. With the rapid development of high-throughput technologies, enormous amounts of biomedical data are being generated and continue to accumulate exponentially. Today, it is even possible to extract the DNA sequence information from individual cells [1], generating millions of data points from a single cell. This vast repository of invaluable data together held in biomedical repositories (covering entities such as diseases, drugs, genes, pathways, etc.) have great potential to uncover new therapeutic options for patients. The key challenge, however, is to make sense out of this, taking into account the volume and variety of available data, and the complex relationships connecting them.

As a computational biopharma company, NuMedii has access to millions of high quality data on diseases and drugs at different stages of development. We built a semantic Knowledge Base compliant with W3C standards[2] to integrate and unify heterogeneous data from various public and proprietary data sources; and to build meaningful relationships between them. This Knowledge Base empowers us to ask questions pertaining to multiple data sources, by executing a single SPARQL query. We use OntoText GraphDB[3], a highly scalable RDF graph database, which includes triple store, inference engine and SPARQL query engine. In addition, we developed a graphical user interface, which enables the domain experts to explore the Knowledge Base by interacting with graphical elements (Figure 1).

---

[1] https://www.nature.com/articles/nmeth.2769
[2] https://www.w3.org/standards/
[3] https://ontotext.com/products/graphdb/

2   Case Study

Idiopathic Pulmonary Fibrosis (IPF) is a lung disease with unknown origin, which causes fibrosis (scarring) in the lungs. IPF is a heterogeneous disease, i.e. mixture of cell types are involved. Current treatments, *pirfenidone* and *nintedanib* [4], slow the fibrosis progression at best, but there is no cure for IPF.

To accelerate the identification of effective drugs for IPF, we built a Knowledge Base (Section 1), that helps us traverse relevant information about IPF very rapidly. Since IPF is a fibrotic disease, we also included data of other fibrotic diseases. This has resulted in a unique and unified resource of Fibrotic diseases which so far contains around 8 billion explicit and inferred RDF statements. Figure 1 shows a screenshot from the user interface visualizing requested data from the Knowledge Base. In addition, we mined more than 700,000 PubMed abstracts on different fibrotic diseases, RDFized the results and added them to the Knowledge Base.
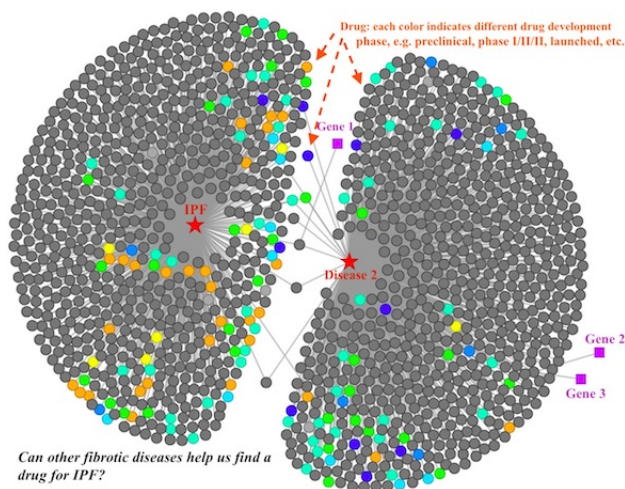

Fig. 1: Screenshot from the user interface shows drugs developed (in any phase) for IPF and another fibrotic disease; as well as genes targeted by the selected drugs.

3   Conclusion

As a critical step towards finding effective drugs for incurable diseases such as IPF, NuMedii adopted Semantic Web technologies to integrate and interrogate all relevant data in one place in a unified fashion; and to infer new knowledge. The interactive graphical user interface allows scientists explore the Knowledge Base without knowing RDF and SPARQL. Moreover, To address the challenges of large graph visualization, the interface allows to tailor the graphs and shed light on nodes and relationships of interest. Furthermore, the Knowledge Base empowers NuMedii to evaluate the drugs predicted by the company's proprietary algorithms. Both the Knowledge Base and the user interface are expandable and have been applied by NuMedii to other complex disease types.

---

[4] https://www.drugbank.ca/drugs/DB04951 and https://www.drugbank.ca/drugs/DB09079