

Title: PCaKG – a prostate cancer knowledge graph

Authors:

Tuulikki Suomela, Xin Lai

Keywords:

cancer, bioinformatics and computational biology, prostate cancer, knowledge graph

Abstract

Prostate cancer is the second leading cause of cancer deaths in men. Prostate cancer is a heterogenous disease, the development of which involves complex genetic alterations and dysregulation in signalling pathways. Exploring these complex networks could offer insights into prostate cancer biology and aid in the development of new therapies.

Knowledge graphs, which represent information as networks of interconnected entities and their relationships, have been adopted to explore cancer data. Compared to traditional tabular databases, knowledge graphs are more efficient in exploring highly interconnected biological data and more flexible in incorporating new types of entities and relationships.

We have developed PCaKG, a knowledge graph for prostate cancer, which aims to comprehensively model the genetic alterations, molecular interactions, clinical characteristics, pathways and ontologies involved in prostate cancer by incorporating publicly available data. PCaKG incorporates clinical and multi-omics data from 343 patients, as well as single-cell RNA-seq data from 19 patients. PCaKG represents prostate cancer using 19 types of entities, over 100 000 entities, 54 types of relationships and more than 23 million relationships. To interpret the vast information in PCaKG, the biomedical AI agent Biomni has been successfully utilised to answer prostate cancer questions presented in natural language based on information from PCaKG.

PCaKG demonstrates the potential of knowledge graphs in efficiently exploring complex cancer networks. When integrated with an AI model, PCaKG could aid in developing more robust and explainable AI models for answering prostate cancer questions and making clinically useful predictions.