

# Roadmap for a Cancer Knowledge Graph

Leily Rabbani, Karolinska Institutet

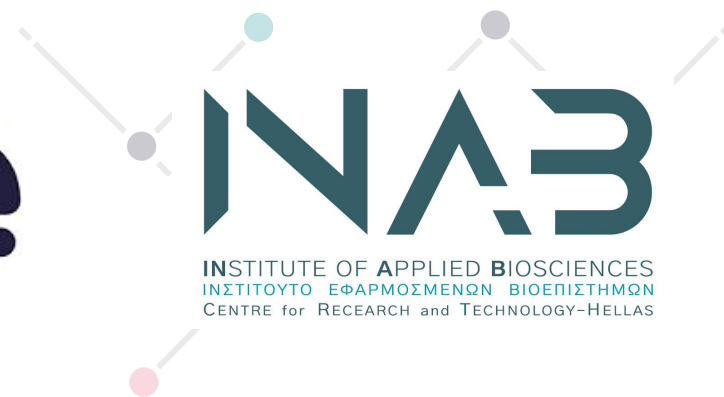
OSFAIR 2023 - Workshop on Open Science Knowledge Graphs

26<sup>th</sup> Sept 2023



**Karolinska  
Institutet**

**Sci!ake**



INSTITUTE OF APPLIED BIOSCIENCES  
ΙΝΣΤΙΤΟΥΤΟ ΕΦΑΡΜΟΣΜΕΝΩΝ ΒΙΟΕΠΙΣΤΗΜΩΝ  
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

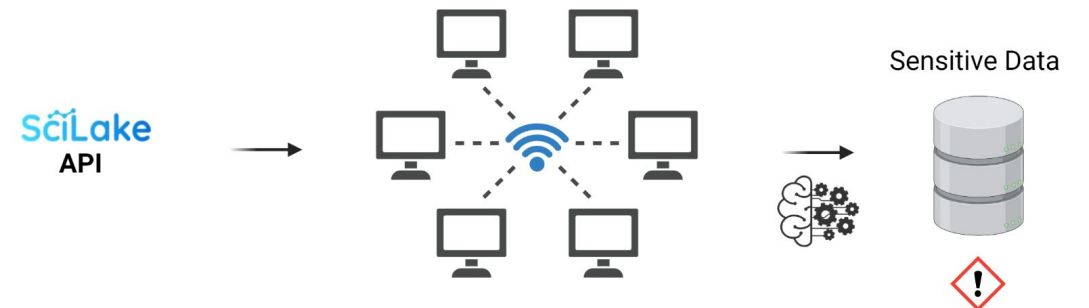
# Introduction

- Create a CLL-specific KG using Scilake functionalities
- Contextualize and connect information
- Inspire follow-up experiments
- Be used as an input for ML tools/libraries

## 1 Patient Stratification



## 2 Federated Machine Learning



# What is Chronic Lymphocytic Leukemia (CLL)

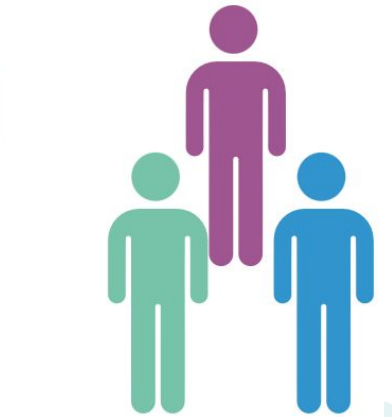
- Characterized by the accumulation of neoplastic B-cells in the bone marrow, peripheral blood & secondary lymph nodes
- Heterogeneous genetic landscape
- Heterogeneous clinical outcome
- The most common adult leukemia
- Still incurable



Initiation



Progression



Drug response



# SciLake Cancer Research Pilot

- Involved partners:

- Karolinska Institutet
- INAB | CERTH

- Target Audience:

- Researchers
- Clinicians

- Objectives:

- Being able to address ongoing research questions such as
  - “Why are particular mutations predictive of overall survival?”
- Better understanding of molecular biology and immunopathology of CLL
- Studying the the possible effect of the mutations



# Benefits & Challenges

## *Benefits*

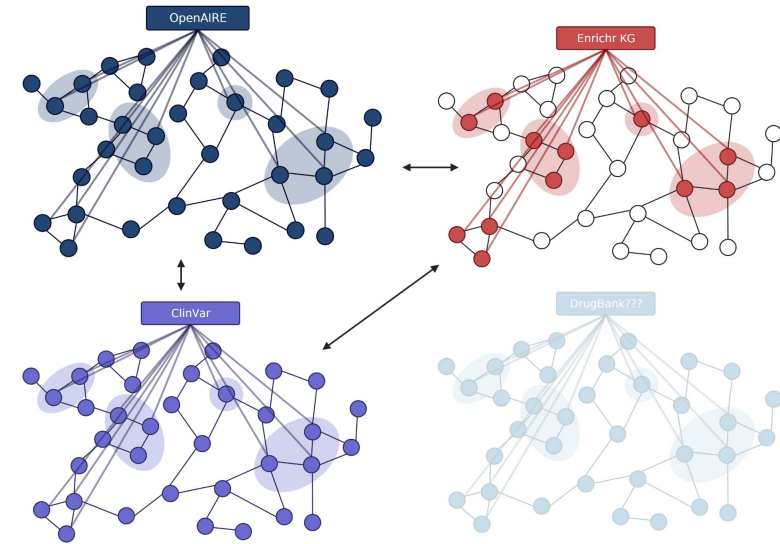
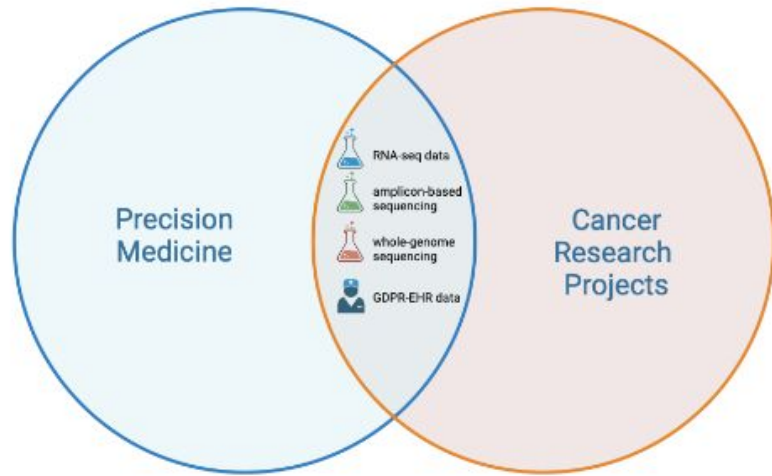
- Enriching the patient-specific data
- Inferring potential associations of the identified biomarkers to additional elements (genes, pathways, drugs, etc.)
- Filtering the graph for the criteria of interest
- Data Integration from various sources
- Possibility of updating the KG

## *Challenges*

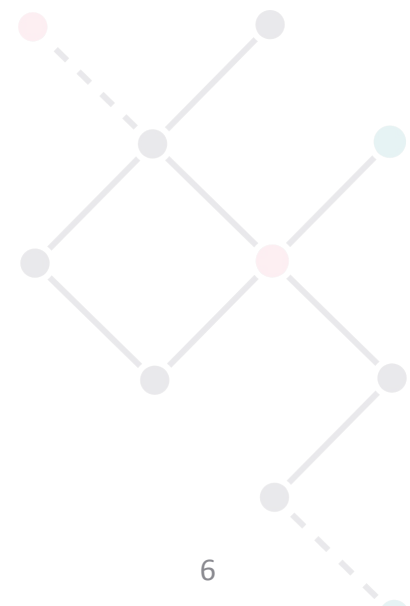
- Validating the KG
- Learning new connections in the graph



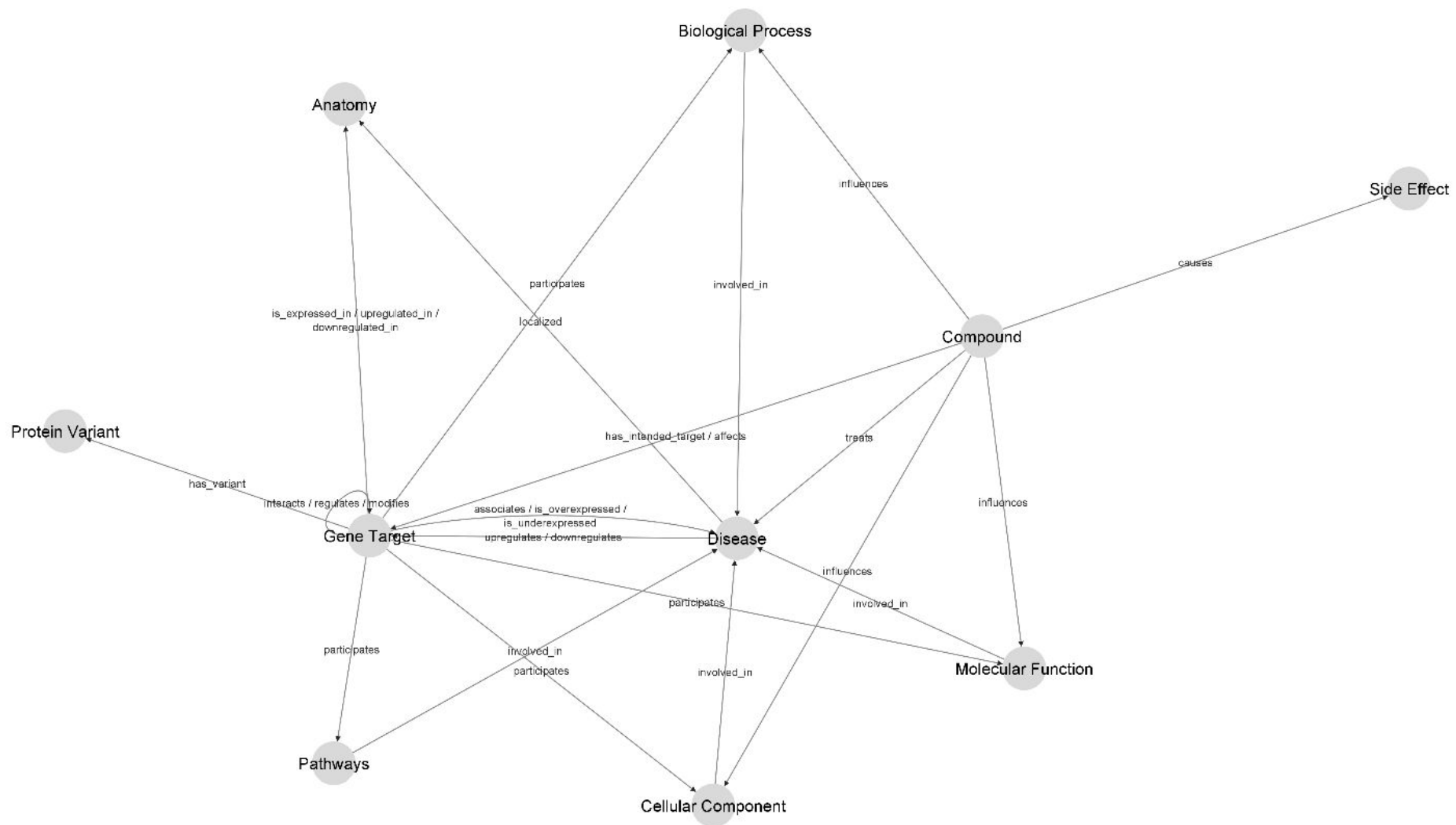
# Dataflow towards a CLL KG



Aggregated output



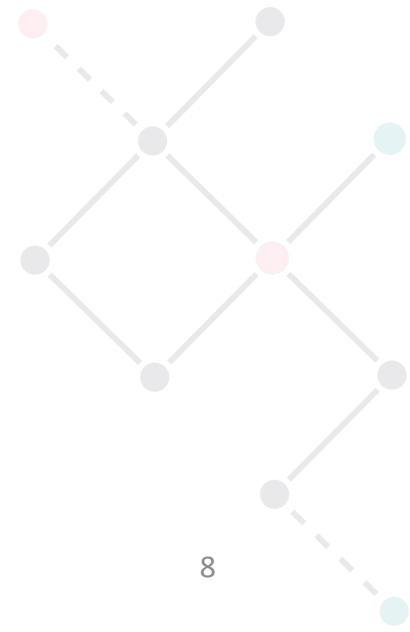
# Desired outlook





# Expected Outcomes

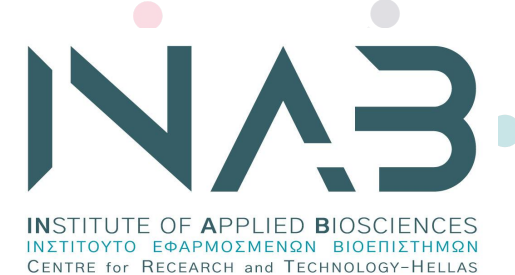
- Using of the Scientific Lake service to construct a domain-specific KG
- Addressing ongoing research questions such as patient stratification
- Taking advantage of the impact and reproducibility analysis service to prioritise findings
- Better understanding of molecular biology and immunopathology of CLL
- Studying the the possible effect of the mutations
- Utilising multiple sources of information







**Karolinska  
Institutet**



**Thank you!  
Questions?**



**Sci!ake**

