LinkedImm: A Linked Data Approach for System Immunology

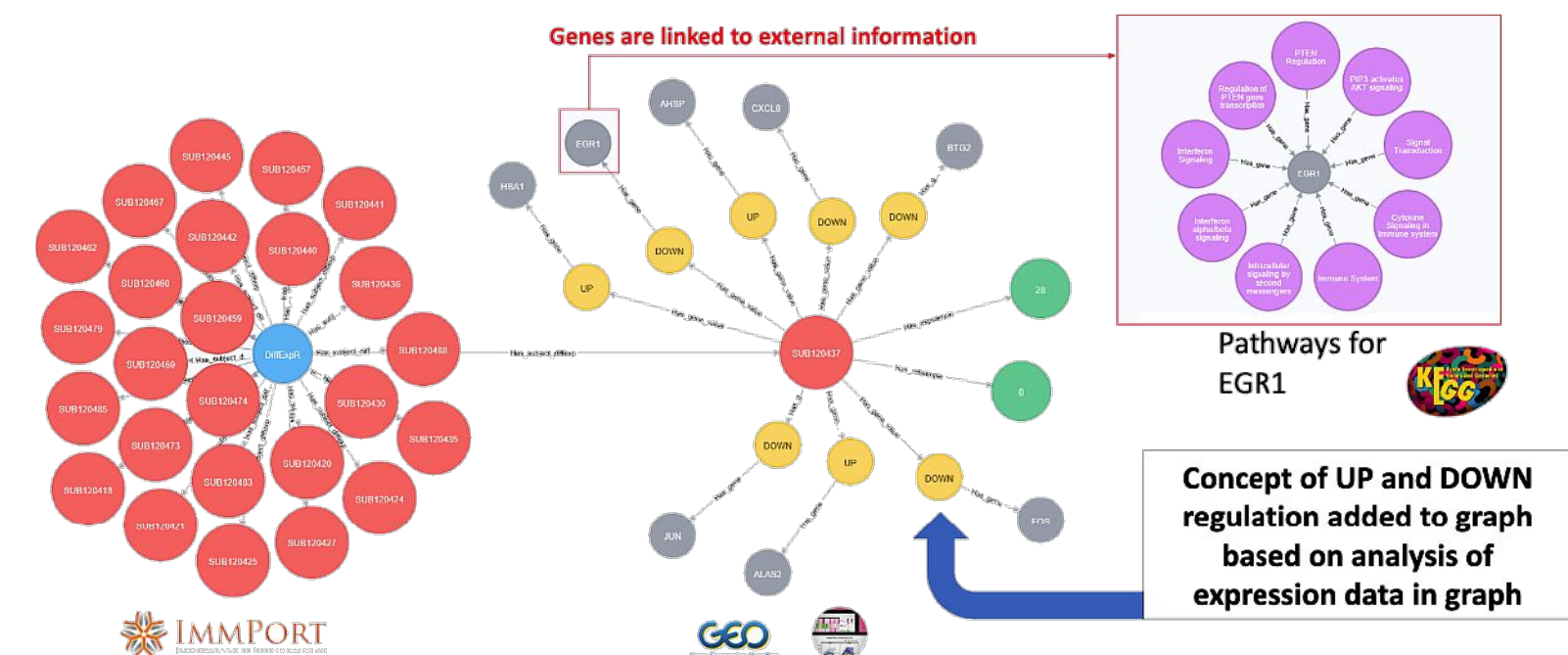Syed Ahmad Chan Bukhari[1], Kei-Hoi Cheung[3,4] and Steven H. Kleinstein[1,2]

[1]Department of Pathology, Yale School of Medicine, New Haven, CT, USA, [2]Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, USA, [3]Department of Emergency Medicine, [4]Yale Center for Medical Informatics, Yale School of Medicine, New Haven, CT, USA.

Systems vaccinology studies combine high-throughput experimental profiling techniques to provide an integrated, dynamic view of vaccine-driven immune responses. In this work, we introduce **LinkedImm**, which integrates heterogeneous data generated from profiling the human response to influenza vaccination data along with other public biological datasets. Integration of heterogeneous data is challenging because of their diverse data formats. To facilitate data integration, we designed a data model which logically arranges the key data points such as genes, pathways, transcriptional profiling data available at various public repositories (e.g., Immport, ImmuneSpace, Reactome and Gene Ontology). These heterogeneous data are then made available in a unified format through a knowledge graph. Our ongoing work on LinkedImm dashboard will enable the scientific community to easily explore theses integrated data.
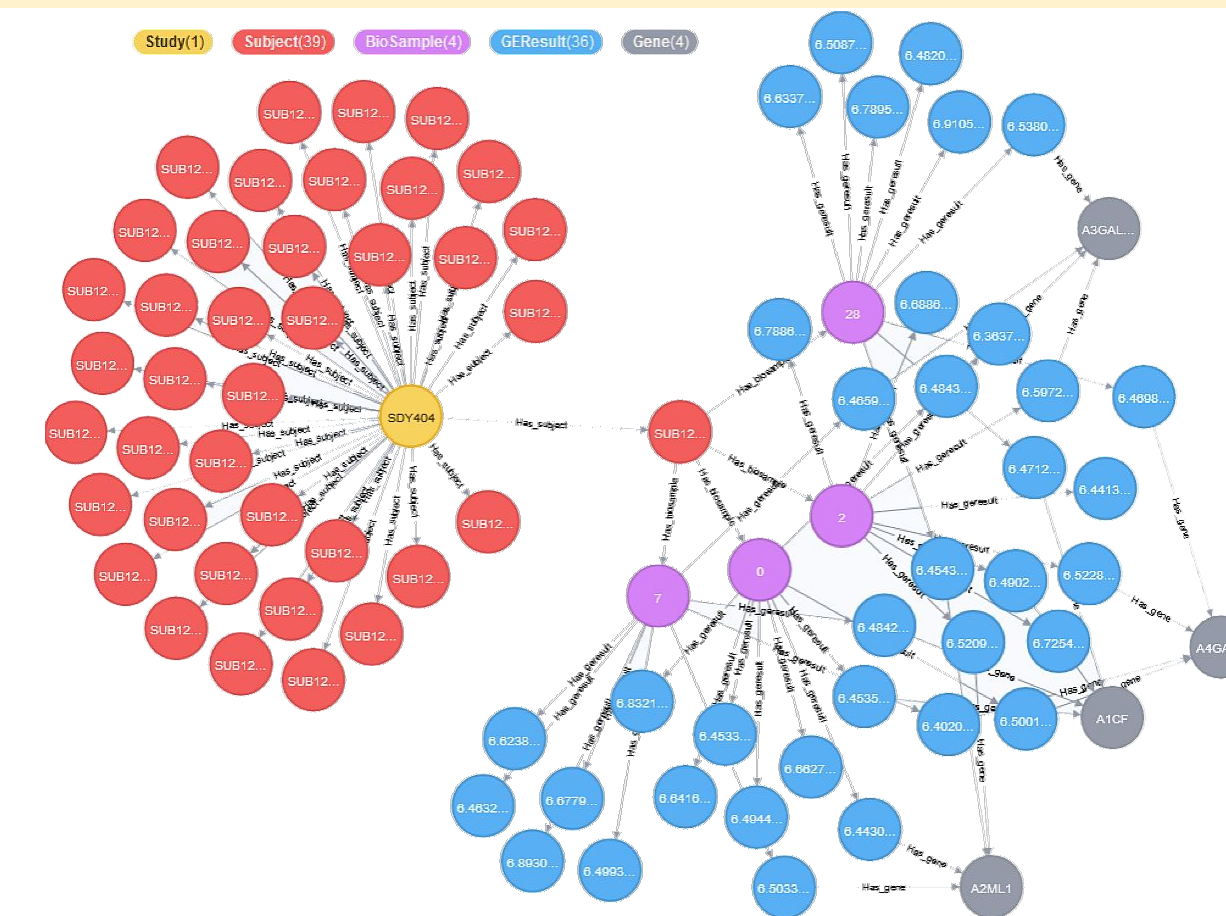
Use Case: Heterogeneous Immunology Data Need Integration

Time (days post-vaccination)

Multiple assays are used to collect various measurements at different time points.

Data Silos

Multiple Formats

The LinkedImm framework consists of:

Data Model: provides extendable schema to logically integrate heterogeneous data.

Knowledge Graph: NoSQL-based data warehouse, used for secondary analysis.

Interface: allow users to explore data through dashboard or natural language.

LinkedImm Flow Diagram

Results : Heterogeneous Data Integration

Genes are linked to external information

Pathways for EGR1

Concept of UP and DOWN regulation added to graph based on analysis of expression data in graph
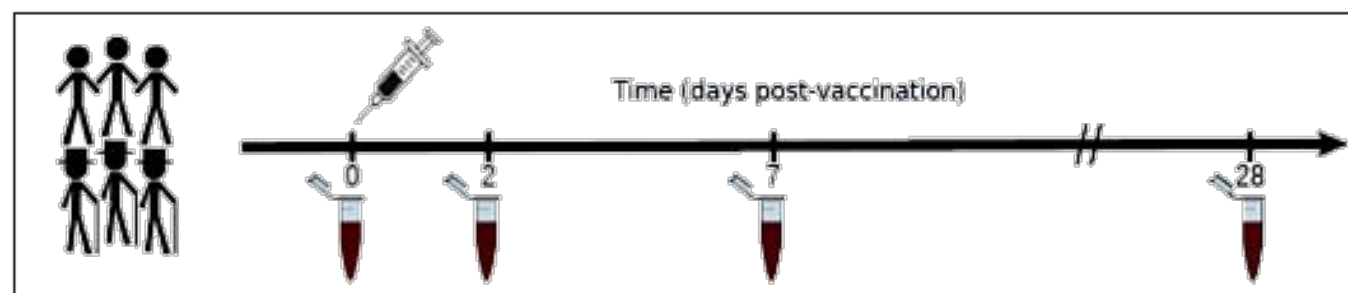
LinkedImm Gene Expression Sub-graph has > 17.5K genes.
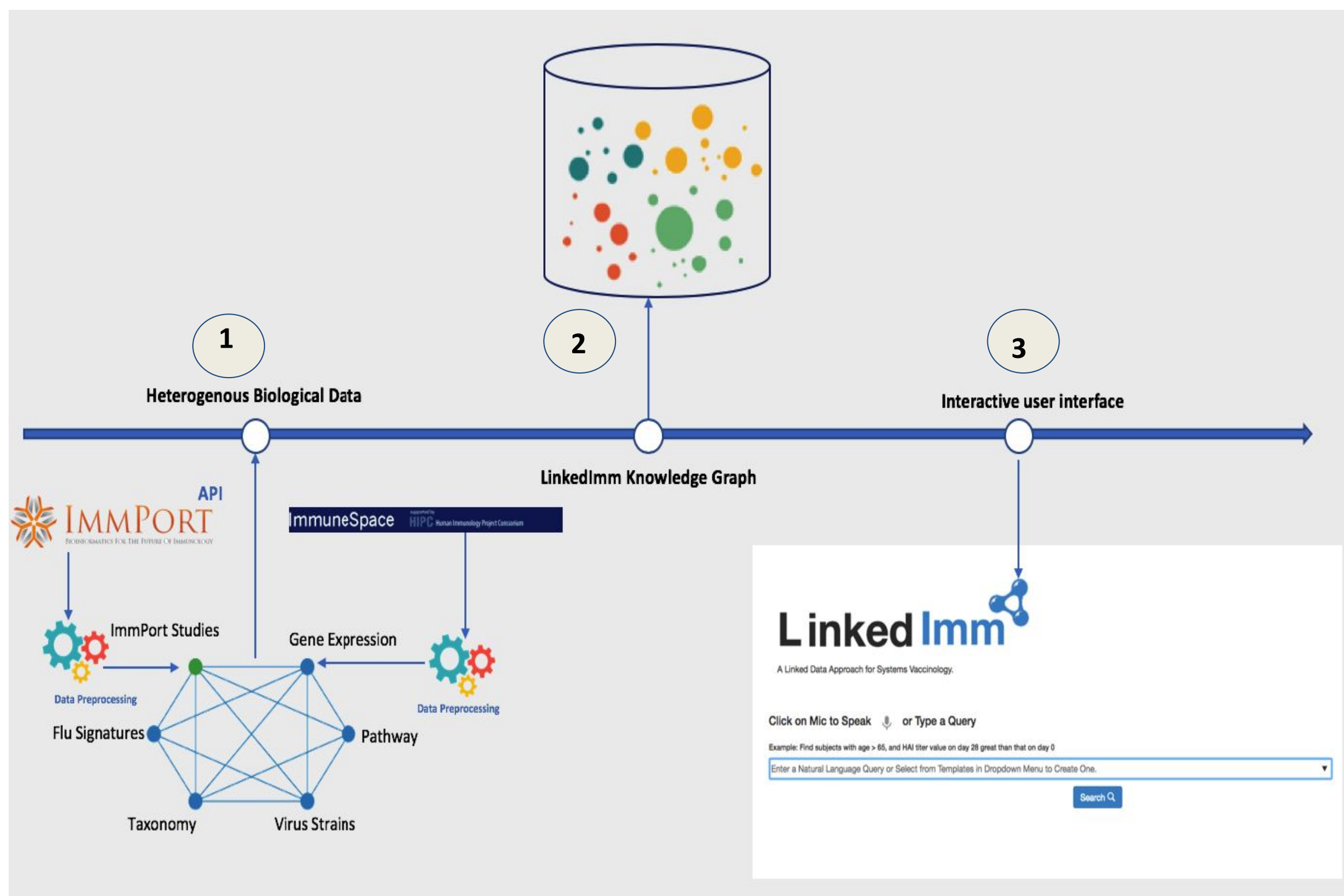

LinkedImm Summary

- ❏ Integrates four different data formats (relational database, ontologies, spreadsheet and graph data).
- ❏ 54 influenza vaccination studies with demographics and HAI results pre- and post-vaccination.
- ❏ 2,256 subjects + 22,525 virus strain measurements.