SPOKE: Connecting Biomedical Information on Earth and in Space via the SPOKE Knowledge Graph

Sergio Baranzini, University of California, San Francisco

Use case description and societal challenge being addressed

SPOKE is a biomedical knowledge graph which currently contains 35 million nodes of 23 different types and 70 million edges of 75 types using information downloaded from 43 databases. In this project, we propose to direct development of SPOKE towards space biology and medicine to facilitate bi-directional translation of knowledge and identification of meaningful analogues to maximize ongoing efforts to make humans a multiplanetary species. Specifically, the *main goal* of this proposal is to engineer the SPOKE knowledge graph to interface with NASA's GeneLab platform and enable seamless integration of primary experimental data across domains of interest for space biology. *A second aim* will focus on the integration of *social determinants of health* (SDOH) into SPOKE, to link these important health modifiers with molecular, cellular, and pathophysiological processes in a data-driven manner and ultimately increase its societal impact.

Researchers can gain unprecedented access to a wealth of multi-species biological information and leverage it to make novel discoveries and advance scientific knowledge. By integrating data from diverse experiments, scientists can identify shared patterns, uncover novel gene-drug interactions, and investigate the effects of space-related stressors on gene expression and molecular pathways, ultimately aiding in the development of targeted countermeasures to mitigate health risks. This knowledge will not only enhance our understanding of space biology but also contribute to broader areas of genetics and genomics research. The project will encourage interdisciplinary collaboration among researchers from various scientific disciplines. This collaborative effort will foster synergies between disciplines and promote the exchange of expertise, leading to novel approaches for studying space biology and its applications beyond the scope of spaceflight.

The integration of biological and SDOH has direct implications for space exploration. The findings of this project can contribute to the development of personalized countermeasures, tailored healthcare interventions, and improved astronaut well-being. These advancements have the potential to enhance the success and safety of long-duration space missions, advancing the frontiers of human exploration beyond Earth.

Knowledge graph source datasets

Datasets for this project come from NASA's GeneLab (e.g. RNA expression in model organisms, microbiome profiling, miRNA, etc), EPA (superfunds, air quality, etc), CDC (disease prevalence by location), WHO (SDOH, disease prevalence, etc), US Census data (SDOH). In this project, we expect to incorporate hundreds of thousands new nodes, and up to a million new relationships into SPOKE.

User queries / competency queries for the use case

Examples of user queries:

- What conditions can be inferred from the gene activity pattern of blood cells in space
- What drugs can counteract or minimize the effect of radiation in lower orbit and in outer space
- What are the most physiologically accurate Earth analogs of spaceflight?
- What conditions on Earth most closely resemble long spaceflight missions?
- How are food deserts associated with diabetes?
- Where is air pollution most closely associated with lung cancer in non-smokers?