U01: Benos and Sciurba - Systems level causal discovery in heterogeneous TOPMed data

[Introduction slides from 6/16/17 Analysis Committee call](#) [1] **(log-in required)**

The advent of new technologies for collecting and analyzing multiple heterogeneous data streams from the same individual makes possible the detailed phenotypic characterization of diseases and paves the way for the development of individualized precision therapies. A major bottleneck in this process is the lack of robust, efficient and truly integrative analytic methods for such multi-modal data. This proposal builds on the ongoing efforts of our group in the area of causal learning in biomedicine. The objective of this application is to extend, modify and tailor our causal probabilistic graphical models to data typically collected by TOPMed projects, such as –omics data (SNPs, metabolomics, RNA-seq, etc), imaging, patients' history, and clinical data.

COPDGene® is one of the TOPMed projects and has generated datasets with those modalities for 10,000 patients with chronic obstructive pulmonary disease (COPD), the third leading cause of death and a major cause of disability and health care costs in the US. The prevailing view is that COPD is a syndrome, consisting of multiple diseases with different characteristics. There is currently no satisfactory method for COPD subtyping or prediction of disease progression. In this project we will apply, test and validate our approaches on COPDGene® and another large independent COPD cohort. The extension and application of our methods to cross-sectional and longitudinal data will also allow us to investigate a number of important questions and aspects related to COPD.

Mechanistically, we will investigate how SNPs, genes and their networks are causally linked to disease phenotypes. In pathology, we will identify conditional biomarkers, which will lead to disease sub-classification and identification of causal components in each subtype. In pathophysiology, we will identify features that are directly linked to lung function decline and outcome.

We will make all our algorithms and results available to the community through web and public cloud interfaces. The deliverables will be (1) new probabilistic approaches for integration and analysis of multi-modal cross-sectional and longitudinal data, including SNPs, blood biomarkers, CT scans and clinical data; (2) new cloud-based server to make these approaches available to the research community; (3) results on the mechanism, pathology and pathophysiology of COPD facilitation and progression. To guarantee the success of the project we have assembled a team of experts in genomics, machine learning, cloud computing and COPD. This cross-disciplinary team project will have a positive impact beyond the above deliverables, since the generality of our approaches makes them applicable to any disease.

We expect that during this U01 we will have the opportunity to collaborate with other teams in the TOPMed consortium to help them investigate the causes of their corresponding disease phenotypes. We do believe that data integration in a single probabilistic framework will be in the heart of precision medicine strategies in the future, when massive high-throughput data collection will become a routine diagnostic and prognostic procedure in all hospitals.

NHLBI Program officer: Maarten Leerkes

PI:	Takis Benos
Co-PI:	Frank Sciurba
Award Type:	U01 NHLBI TOPMed Program: Integrative Omics Approaches for Analysis of TOPMed Data (RFA-HL-17-011)
Award number:	U01 HL137159-01
Start Year:	2017