U01: Kang - Scalable and translational analysis tools on the cloud for deep integrative omics data

The NHLBI Trans-Omics for Precision Medicine (TOPMed) program aims to provide high- priority studies of heart, lung, blood and sleep disorders (HLBS) with high-quality genomic data. This year, the program will deeply sequence >60,000 genomes to characterize DNA sequence variation at scale. It is expected that >400 million genetic variants will be identified. In later phases, it is expected that rich genomic assays will be applied to an equally large number of samples. In a pilot phase, these additional assays will include ~3,000 transcriptomes, ~2,000 methylation profiles, and ~2,000 metabolomics profiles.

Data on this scale opens up many opportunities for discovery and analysis but also poses significant challenges. RFA-HL-17-011, entitled "NHLBI TOPMed Program: Integrative Omics Approaches for Analysis of TOPMed Data (U01)" is intended to stimulate development of computational and statistical methods and tools that enable innovative and scalable analyses genomic resource. Our group has a long history in the development of specialized, state-of-the- art methods and tools for the processing and analysis of large genomic datasets. We have a history of leadership in varied resources, ranging from the Mouse HapMap Project, to 1000 Genomes Project, to ENCODE, and including the NHLBI's TOPMed program. In this application, we propose to develop innovative and practical methods to enable informative genomic analysis at scale.

These methods encompass computational tools to rapidly scale deep GWAS, statistical methods for robust and powerful integrative omics analysis, and visualization methods for integrative interpretation of omics genetics results. We will implement these methods into cost-effective, easy-to-use, and well-documented software packages that facilitate understanding of molecular mechanisms involved in HLBS disorders. A key component of the proposal is the deployment of these tools on commercial clouds, providing accessible interface to investigators without direct access to a local high-throughput compute and data storage facility. The resulting tools will empower a wide range of scientists to run best-in-class methods to accelerate discovery of new treatments for HLBS disorders.

_____

NHLBI Program officer: Rebecca Beer

PI:	Hyun Min Kang
Award Type:	U01 NHLBI TOPMed Program: Integrative Omics Approaches for Analysis of TOPMed Data (RFA-HL-17-011)
Award number:	U01 HL137182-01
Start Year:	2017