# Transcriptomics-based Machine Learning (ML) Analysis Predicts Space-Exposed Murine Livers

Hari Ilangovan[1], Prachi Kothiyal[2], Katherine A. Hoadley[3], Robin Elgart[4], Greg Eley[2], Parastou Eslami[5]
[1]Science Applications International Corporation (SAIC), Reston, VA 20190, USA
[2]Scimentis LLC, Statham, GA 30666, USA
[3] Department of Genetics, Computational Medicine Program, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[4]University of Houston, Houston, TX 77204, USA
[5]Universal Artificial Intelligence Inc, Boston, MA 02130, USA

NASA has employed high-throughput molecular assays to identify sub-cellular changes impacting human physiology during spaceflight. Machine learning (ML) methods hold the promise to improve our ability to identify important signals within highly dimensional molecular data. However, the inherent limitation of study subject numbers within a spaceflight mission minimizes the utility of ML approaches. To overcome the sample power limitations, data from multiple spaceflight missions must be aggregated while appropriately addressing intra- and inter-study variabilities. Here we describe an approach to log transform, scale and normalize data from six heterogeneous, mouse liver derived transcriptomics datasets ($n_{total}$=137) which enabled ML-methods to perform well (AUC ≥ 0.87) in classifying spaceflown vs ground control animals rather than mission-of-origin. Concordance was found between liver-specific biological processes identified from harmonized ML-based analysis and study-by-study classical omics analysis. This work demonstrates the feasibility of applying ML methods on integrated, heterogeneous datasets of small sample size.